Comparative effectiveness of glycemic control in patients with type 2 diabetes treated with GLP-1 receptor agonists: a network meta-analysis of placebo-controlled and active-comparator trials

Michelle E Orme 1
Hiep Nguyen 2
Jackie Y Lu 3
Susan A Thomas 3

1ICERA Consulting Ltd, Swindon, UK;
2AstraZeneca, Wilmington, DE;
3AstraZeneca, Fort Washington, PA, USA

Background: Clinical studies of patients with type 2 diabetes show that GLP-1 receptor agonists (GLP-1 RAs) improve glycemic control and promote weight loss. We conducted a Bayesian network meta-analysis (NMA) of placebo- and active-controlled randomized trials to assess the comparative effectiveness of liraglutide, albiglutide, dulaglutide, and exenatide twice daily and once weekly, with a focus on glycemic control.

Materials and methods: We searched Medline, Embase, and the Cochrane Library (up to December 2014) for core registration programs for US-approved GLP-1 RAs. Patients reaching an A1C target of <7% were analyzed with a binomial model and change in A1C from baseline with a normal model. A covariate analysis assessed the impact of baseline A1C and treatment background on outcomes.

Results: The base-case NMA used 23 trials reporting A1C outcomes at ~6 month follow-up. The results, unadjusted and adjusted for baseline A1C, indicated that all GLP-1 RAs resulted in statistically significantly lower A1C at follow-up compared with placebo. The odds of reaching the <7% target were also significantly better compared with placebo. With dulaglutide, exenatide once weekly, and liraglutide, the absolute reduction in A1C at 6 months was 0.9%–1.4%, and was significantly better than exenatide twice daily. Albiglutide was not significantly different from exenatide twice daily. We estimate that ~50% of patients will meet the <7% A1C target within 6 months of commencing GLP-1 RAs.

Conclusion: This was a comprehensive assessment of the comparative effectiveness of GLP-1 RAs and A1C outcome. GLP-1 RAs are a viable addition to oral antidiabetes therapy, and dulaglutide, exenatide once weekly, and liraglutide are the most effective.

Keywords: type 2 diabetes, glucagon-like peptide-1-receptor agonists, GLP-1 RAs, network meta-analysis, comparative effectiveness

Background
Comparative effectiveness research (CER) in diabetes is increasingly important, given the wide range of treatment options available to patients with type 2 diabetes (T2D). Over the years, the goals of diabetes management have expanded beyond glycemic control to include the management of metabolic and cardiovascular comorbidities according to several international guidelines.1–5 Several newer classes of antihyperglycemic agents, including GLP-1 RAs and sodium–glucose cotransporter inhibitors, have been suggested to provide additional benefits, such as weight loss. Both the American Association of Clinical Endocrinologists guidelines and the American Diabetes...
Association recommend a patient-centered approach to guide choice of pharmacological agents.\textsuperscript{2,6} Considerations include efficacy, cost, potential side effects, weight, comorbidities, hypoglycemia risk, and patient preferences. Both bodies recognize that GLP-1 RAs have robust $\text{A}_{1c}$-lowering properties, are usually associated with weight loss and blood-pressure reductions, and are available in several formulations. The risk of hypoglycemia with GLP-1 RAs is low, and they reduce fluctuations in both fasting and postprandial states.

GLP-1 RAs are a growing class of glucose-lowering drugs that improve glucose homeostasis by enhancing the endogenous secretion of insulin induced by meal ingestion, inhibiting glucagon secretion, and slowing gastric emptying. Notably, they also suppress food intake and appetite, through central effects.\textsuperscript{7} Since the first GLP-1 RA was approved in 2005, the number of injectable agents in this class has increased from exenatide twice daily (EBID [exenatide bis in die]) to include liraglutide (Lira) once daily, exenatide once weekly (EQW [exenatide quaque week]), albiglutide (Albi) QW, and dulaglutide (Dula) QW.

Given the wide choices of GLP-1 RA agents, CER can be a useful tool to aid health care decision makers weigh up the benefits and harms associated with different treatment options. A common CER approach is to synthesize the available randomized controlled trial (RCT) evidence in a meta-analysis to provide a comprehensive view of the relative efficacy of the treatment options. The standard direct meta-analysis method is limited to evaluating the relative efficacy of treatments in a pairwise manner, where all the trials included in the direct meta-analysis compare the same intervention with the same control. Many trials are either placebo-controlled, include an active control that does not represent the current standard of care, or may not be comparable to the active arm in a treatment decision-making context. In the absence of head-to-head trials, indirect comparisons can be made using a common control arm to bridge the gap, provided that the randomized comparisons within each trial are preserved.\textsuperscript{8,9} Network meta-analysis (NMA), an extension of the standard meta-analysis methods, calculates the relative effects for all treatments in the evidence network in one simultaneous analysis.\textsuperscript{10–12} NMA is different from pairwise meta-analysis in the sense that there is not only one type of treatment comparison, but multiple treatment comparisons. Therefore, the NMA output provides a comprehensive evidence base that allows decision makers to compare the effects from any two treatments within the network, including the relative-effect estimates between treatments that have not been compared in head-to-head trials. NMAs can also provide more precise estimates of treatment differences than can be obtained from pairwise meta-analysis, since more of the data are used.\textsuperscript{10–12}

There are several published meta-analyses evaluating the clinical profile of GLP-1 RAs;\textsuperscript{13–19} however, these analyses either had limited data for more recent US Food and Drug Administration-approved GLP-1 RAs, including EQW formulations, did not apply the NMA methods to compare the relative efficacy of GLP-1 RAs, used a frequentist NMA method, or did not control for baseline $\text{A}_{1c}$. Therefore, we performed a Bayesian NMA of placebo-controlled and active-controlled randomized trials to assess the relative effect of Lira, Albi, Dula, EBID, and EQW, with a particular focus on glycemic control.

Materials and methods

Study selection and data sources

We identified eligible studies by searching Medline, Embase, and the Cochrane Library from inception up to December 31, 2014, using pertinent keywords, and restricted our results to published RCTs in the English language. We included RCTs from core registration programs for all US-approved GLP-1 RAs. We included open-label and double-blind RCTs comparing one GLP-1 RA with another, at any dose or with a control (placebo, oral antihyperglycemic drugs), for adults with T2D. For inclusion, studies had to fulfill the following criteria: 1) placebo-controlled or active-comparator RCTs comparing one GLP-1 RA with another, at any dose or with a control (placebo, oral antihyperglycemic drugs), for adults with T2D; 2) reported outcome of percentage of patients achieving $\text{A}_{1c} < 7\%$ baseline; 3) provided mean change in $\text{A}_{1c}$ from baseline with standard error or 95\% confidence interval [CI]; and 4) to be included in the base-case analysis, outcomes needed to be reported at 6-month follow-up (included range 24–32 weeks).

Data extraction

Data were from full-text publications for all GLP-1 RA RCTs used for product registration in the US. Data were extracted by two reviewers, and discrepancies resolved by consensus. For some studies, it was necessary to supplement the data extraction with information from clinical trial-registration records.\textsuperscript{20}

Outcome measurements and treatments

The outcome data extracted included the percentage of patients with $\text{A}_{1c}$ below target of 7\% at follow-up and change from baseline in $\text{A}_{1c}$ at follow-up. Change in $\text{A}_{1c}$ from baseline was most commonly reported on a modified
intent-to-treat basis where the population was defined as the set of patients who were randomized, received at least one dose of study medication, and had at least one postbaseline A1C measurement. For the A1C target outcome, trials reported the percentage of patients reaching treatment targets on an evaluable case basis, whereby only the subset of patients with A1C >7% at baseline was evaluated.

For the NMA, we pooled data by GLP-1 RA agent: Albi, Dula, EBID, EQW, and Lira. GLP-1 RA treatment arms at unlicensed doses were not included. Other control arms were included in the evidence networks to preserve randomization, and these were pooled by treatment class as follows: placebo, DPP4 inhibitor, insulin, metformin, sulfonylurea, and thiazolidinedione.

**Meta-analyses**

The NMA approach was as per the UK’s National Institute for Health and Care Excellence (NICE) Decision Support Unit recommendations for Bayesian NMA. This methodology is widely used for synthesizing clinical trial data for health-technology appraisal or regulatory purposes. The Bayesian statistical model applies Monte Carlo simulations, which converge the direct (A versus B) and indirect (A versus C, C versus B) evidence with the likelihood-effect estimate, and provides a modeled comparison between A versus B versus C. The underlying assumption of this approach is that the comparator group for the interventions (ie, C) is similar among the indirect-comparison trials. Continuous outcomes were analyzed using a normal model with an identity link, and dichotomous outcomes were analyzed using a binomial model with logit link. Both fixed-effect and random-effect models were investigated. Fixed and random-effect models were fitted to the data via Bayesian Markov chain Monte Carlo methods using WinBUGs 1.4 and were run in for a minimum of 100,000 iterations to ensure convergence. Subsequently, two chains of 100,000 were sampled from the posterior distributions. These samples were used to calculate the median/mean and the 95% credible interval (CrI), which is the interval from the percentiles 2.5 to –97.5. The CrI, distinct from the CI, is the Bayesian equivalent of the frequentist 95% CI, and is used to assess statistically significant differences, which is consistent with the approach used by NICE in evaluating effectiveness data.

All results for the NMA are reported as medians with corresponding 95% CrIs. Medians are presented as the best estimate for the central value, since means may be overly influenced by outliers. The pooled summary measure for continuous end points is weighted mean differences and odd ratios for binomial outcomes. An estimate of how well the predicted values fitted the observed data set was provided by the mean residual deviances (total residual deviance divided by number of data points), as well as the deviance information criteria (DIC) output from WinBUGs. Models with a good fit would have a total residual deviance close to the number of data points. The DIC is used to compare different models for the same likelihood and data, and the model with the lowest DIC was deemed to best predict a replicate data set of the same structure to that observed. There were no major differences in DIC when comparing fixed-effect with random-effect models. Fixed-effect models assume that differences across trials do not impact on the treatment effects, and that variation in the outcomes reported are due to differences between patients within a trial. Random-effect models assume that variation in the outcomes reported are due both to differences between patients within a trial and differences across trials. Therefore, results from the random-effect NMA models have been presented in this paper, since these better take into account sources of uncertainty.

Covariate analyses were conducted to explore the effect of baseline A1C and use of background treatment that may confound the A1C end point. Previous meta-analyses have shown that there is a correlation between baseline A1C and change in A1C over follow-up. Therefore, a continuous study-arm level variable for baseline A1C was included in the model, centered at the mean baseline A1C across all study arms, the assumption being that the baseline A1C has the same impact on effects across all treatments. In a further covariate analysis, a continuous study-arm level variable for percentage of patients on oral therapy as background (0–100%) and a dummy-indicator variable for use of insulin as background treatment (1, insulin included in background; 0, insulin not used as background treatment) were included to account for differences in background treatment. Note that covariate meta-analysis adjusts for differences between study arms, and as aggregated data are used in the covariate meta-analysis, the results should not be used to make predictions about individual patients.

The majority of studies reported outcomes at approximately 6-month follow-up. Some studies were of a longer duration, and a few studies did not provide sufficient endpoint data at 6-month follow-up in either the full-text publication or the clinical trial-registry record. For the base-case analysis, we used all studies that reported the outcomes of interest between 24 and 32 weeks of follow-up. A sensitivity analysis was conducted using all studies regardless of the follow-up time.
In addition to the NMA, standard direct meta-analysis was also conducted in Stata version 14.33 We pooled studies using fixed- and random-effect models, using the random-effect method of DerSimonian and Laird, where the estimate of between-study heterogeneity is taken from the fixed-effect Mantel–Haenszel or inverse-variance model.30,34 The direct meta-analysis was conducted to supplement the NMA results and to investigate potential inconsistencies between the direct and indirect estimates.

Results
In total, 29 GLP-1 RA core registration trials were identified covering 18,543 patients, which included seven trials for Albi,45–50 six trials for Dula,42–47 four trials for EBID,48–51 six trials for EQW,52–57 and six trials for Lira (Table S1). Seven of these trials provided head-to-head comparisons of GLP-1 RAs (Table 1).41,42,47,52,56,57,63 Given the results of the direct meta-analyses, different inferences can be made depending on which set of head-to-head trials is used, eg, Dula versus EQW via EBID controlled trials or via Lira controlled trials (Table 1). This demonstrates the need for an NMA to provide a comprehensive assessment of the comparative effectiveness of the GLP-1 RAs.

Table 1: Summary of A1C end points from head-to-head GLP-1 RA trials

<table>
<thead>
<tr>
<th>Head-to-head comparison</th>
<th>Study</th>
<th>Reaching A1C target, OR (95% CI)</th>
<th>WMD in A1C, OR (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dula versus EBID</td>
<td>AWARD-142</td>
<td>2.35 (1.73–3.19) Dula better, P&lt;0.01</td>
<td>–0.41 (–0.53 to –0.3) Dula better, P&lt;0.01</td>
</tr>
<tr>
<td>EQW versus EBID</td>
<td>DURATION-152</td>
<td>2.72 (1.91–3.86) EQW better, P&lt;0.01</td>
<td>–0.55 (–0.75 to –0.35) QW better, P&lt;0.01</td>
</tr>
<tr>
<td>Lira versus EBID</td>
<td>LEAD-653</td>
<td>1.57 (1.09–2.27) Lira better, P=0.02</td>
<td>–0.33 (–0.55 to –0.11) Lira better, P&lt;0.01</td>
</tr>
<tr>
<td>Albi versus Lira</td>
<td>HARMONY-761</td>
<td>0.68 (0.52–0.9) Lira better, P=0.01</td>
<td>0.21 (0.08–0.34) Lira better, P=0.01</td>
</tr>
<tr>
<td>Dula versus Lira</td>
<td>AWARD-647</td>
<td>1.02 (0.72–1.44) No difference, P&gt;0.05</td>
<td>–0.06 (–0.2 to –0.08) No difference, P&gt;0.05</td>
</tr>
<tr>
<td>EQW versus Lira</td>
<td>DURATION-657</td>
<td>0.74 (0.57–0.96) Lira better, P=0.02</td>
<td>0.2 (0.06–0.34) Lira better, P&lt;0.01</td>
</tr>
</tbody>
</table>

Note: *Included a placebo arm.

Abbreviations: OR, odds ratio; CI, confidence interval; WMD, weighted mean difference; Dula, dulaglutide; EBID, exenatide bis in die (twice daily); EQW, exenatide quaque week (once weekly); Lira, liraglutide; Albi, albiglutide.

Across all analyses, the random-effect models had a better fit compared with the fixed-effect models in terms of DIC and average residual deviance. The covariate random-effect models, where treatment effects were adjusted for baseline A1C, had a similar fit to the unadjusted random-effect models. While baseline A1C was not a statistically significant predictor of differences in treatment effects, the direction of effect indicates that study arms with a higher baseline A1C will show a larger effect on A1C compared with study arms with lower baseline A1C. Baseline A1C does not appear to be a confounding factor in analysis of the <7% target end point. However, based on model fit and observations from other analyses, the random-effect model adjusted for baseline A1C is likely to provide the most robust results, since this takes into account some of the heterogeneity between studies.

Table 2: Summary of odds ratios of reaching <7% A1C

<table>
<thead>
<tr>
<th>Head-to-head comparison</th>
<th>Study</th>
<th>Odds ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dula versus EBID</td>
<td>AWARD-142</td>
<td>2.35 (1.73–3.19)</td>
</tr>
<tr>
<td>EQW versus EBID</td>
<td>DURATION-152</td>
<td>2.72 (1.91–3.86)</td>
</tr>
<tr>
<td>Lira versus EBID</td>
<td>LEAD-653</td>
<td>1.57 (1.09–2.27)</td>
</tr>
<tr>
<td>Albi versus Lira</td>
<td>HARMONY-761</td>
<td>0.68 (0.52–0.9)</td>
</tr>
<tr>
<td>Dula versus Lira</td>
<td>AWARD-647</td>
<td>1.02 (0.72–1.44)</td>
</tr>
<tr>
<td>EQW versus Lira</td>
<td>DURATION-657</td>
<td>0.74 (0.57–0.96)</td>
</tr>
</tbody>
</table>

Table 3 shows the probability of reaching the target <7% A1C, number needed to treat (NNT), number of patients reaching target per 100 treated, and absolute change in A1C from base case for each treatment group across all analyses. NNT is the number needed to be treated in order to observe one event of interest, and the lower the NNT (closest to 1), the more effective the treatment. We estimate that for every two patients treated with a GLP-1 RA, one will meet the <7% A1C target within 6 months. Patients are more likely to reach the <7% target at 6 months with Dula, EQW, and Lira compared to Albi. For every 100 patients treated, nine more will reach the <7% target if treated with Dula, EQW, and Lira compared to Albi (Table 3).
Sensitivity analysis

The base-case NMA excluded six studies (five studies for Albi and one for Lira) because they had longer follow-ups and did not report A1C outcomes at 6 months. A sensitivity analysis was conducted to examine the impact of inclusion of these six studies on the relative treatment effects of GLP-1 RAs. This resulted in lower treatment effects for Albi compared with the base-case results, although the difference in effect was not significant (Figure 2). The treatment effects for the other GLP-1 RAs were largely unchanged between the 6-month base-case and sensitivity analyses. It was also noted that the inclusion of longer-term studies resulted in a lower treatment effect for sulfonylureas. This result would be due to the inclusion of the HARMONY-3 study, which included a glimepiride
Probability of reaching <7% A1C target and absolute change in A1C at 6 months

<table>
<thead>
<tr>
<th>GLP-1 RA</th>
<th>&lt;7% A1C target (range)</th>
<th>NNT*</th>
<th>NRT per 100 treated</th>
<th>Absolute change in A1C compared with baseline, OR (CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Albi</td>
<td>43.1% (26.4%–61.7%)</td>
<td>2.32</td>
<td>43.1</td>
<td>–1.022 (–1.334 to –0.709)**</td>
</tr>
<tr>
<td>Dula</td>
<td>52.7% (37.8%–67%)</td>
<td>1.9</td>
<td>52.7</td>
<td>–1.178 (–1.404 to –0.952)**</td>
</tr>
<tr>
<td>EBID</td>
<td>36.1% (24%–50.5%)</td>
<td>2.77</td>
<td>36.1</td>
<td>–0.756 (–0.969 to –0.543)**</td>
</tr>
<tr>
<td>EQW</td>
<td>52.1% (36.8%–67.3%)</td>
<td>1.92</td>
<td>52.1</td>
<td>–1.12 (–1.367 to –0.874)**</td>
</tr>
<tr>
<td>Lira</td>
<td>51.2% (36.9%–65.3%)</td>
<td>1.96</td>
<td>51.2</td>
<td>–1.197 (–1.418 to –0.977)**</td>
</tr>
</tbody>
</table>

Notes: *To get one patient meeting target; **significant difference compared to baseline. From random-effect network meta-analysis adjusted for study arm-level A1C at baseline.

Abbreviations: NNT, number needed to treat; NRT, number reaching target; OR, odds ratio; CI, confidence interval; Albi, albiglutide; Dula, dulaglutide; EBID, exenatide bis in die (twice daily); EQW, exenatide quaque week (once weekly); Lira, liraglutide; Glp-1 RA, GLP-1 receptor agonist; DPP4i, DPP4 inhibitor; vs, versus.

arm and reported end points after 104 weeks. Such a finding is consistent with other studies, in that although sulfonylurea treatment can result in a rapid initial response, the effectiveness diminishes over time, resulting in a gradual increase in A1C.64–66 An additional covariate analysis (results not shown) was undertaken to take into account differences in the background treatment across trials. This incorporated two variables: the percentage of patients on an oral antidiabetes drug at baseline

Table 2 Summary of 6-month network meta-analysis results

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Reaching &lt;7% target, OR (95% CrI)</th>
<th>WMD in A1C, (95% CrI)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RE base case</td>
<td>RE adjusted, BL A1C</td>
</tr>
<tr>
<td>Albi vs Pla</td>
<td>3.852 (2.224–6.733)*</td>
<td>3.686 (2.215–6.841)*</td>
</tr>
<tr>
<td>Dula</td>
<td>5.455 (3.97–7.558)*</td>
<td>5.671 (4.073–8.005)*</td>
</tr>
<tr>
<td>EBID</td>
<td>2.882 (2.149–3.942)*</td>
<td>2.886 (2.144–3.947)*</td>
</tr>
<tr>
<td>EQW</td>
<td>5.521 (3.83–8.182)*</td>
<td>5.551 (3.831–8.208)*</td>
</tr>
<tr>
<td>Lira</td>
<td>5.437 (4.063–7.359)*</td>
<td>5.344 (3.956–7.261)*</td>
</tr>
<tr>
<td>Su</td>
<td>4.857 (2.607–9.15)*</td>
<td>4.604 (2.404–8.787)*</td>
</tr>
<tr>
<td>Ins</td>
<td>3.208 (2.04–4.74)*</td>
<td>3.289 (2.245–4.91)*</td>
</tr>
<tr>
<td>Met</td>
<td>3.613 (2.205–6.007)*</td>
<td>3.692 (2.231–6.172)*</td>
</tr>
<tr>
<td>Tzd</td>
<td>3.46 (2.229–5.456)*</td>
<td>3.442 (2.207–5.461)*</td>
</tr>
<tr>
<td>DPP4i</td>
<td>2.134 (1.386–3.318)*</td>
<td>2.163 (1.34–3.385)*</td>
</tr>
</tbody>
</table>

Notes: *Significantly better compared to placebo; **significantly better compared to the active control.

Abbreviations: OR, odds ratio; CI, credible interval; WMD, weighted mean difference; RE, random-effect; BL, baseline; Pla, placebo; Albi, albiglutide; Dula, dulaglutide; EBID, exenatide bis in die (twice daily); EQW, exenatide quaque week (once weekly); Lira, liraglutide; Su, sulfonylurea; Ins, insulin; Met, metformin; Tzd, thiazolidinediones; DPP4i, DPP4 inhibitor; vs, versus.

Table 3 Probability of reaching <7% A1C target and absolute change in A1C at 6 months

 arm and reported end points after 104 weeks. Such a finding is consistent with other studies, in that although sulfonylurea treatment can result in a rapid initial response, the effectiveness diminishes over time, resulting in a gradual increase in A1C.64–66 An additional covariate analysis (results not shown) was undertaken to take into account differences in the background treatment across trials. This incorporated two variables: the percentage of patients on an oral antidiabetes drug at baseline
and the use of insulin as part of the background treatment. Use of a background oral antidiabetes drug was not a significant predictor of treatment effect. There were insufficient studies to assess whether background insulin could have been a potential effect modifier.

Consistency between direct and indirect evidence

When comparing the direct GLP-1 RA head-to-head results shown in Table 1 with the NMA results in Table 2, the results are largely consistent. The direct results comparing Dula, EQW, and Lira versus EBID had corresponding statistically significant results in the random-effect base-case NMA. For the comparisons of Albi, Dula, and EQW versus Lira, the random-effect base-case NMA did not produce statistically significant results, although statistically significant differences ($P<0.05$) were reported in HARMONY-7 (Lira better than Albi) and DURATION-6 (Lira better than EQW). It was noted that both HARMONY-7 and DURATION-6 had an open-label design and unmatched administration of study drugs (daily versus weekly).

In the course of conducting the analysis, it was noted that AW ARD-1 and Moretto et al$^{51}$ reported that the efficacy of 10 μg EBID was not significantly different from placebo (odds ratio versus placebo 1.46, [95% CI 0.94–2.26], and 2.14, [95% CI 0.99–4.63], respectively). As the treatment-effect CI for Moretto et al overlapped the CIs estimated from the other three EBID trials (Figure 3) and given the study size (56 and 59 patients in the exenatide and placebo arms, respectively), we attribute the lack of significance to a lack of power to detect differences between study arms. AW ARD-1, on the other hand, included sufficient numbers of patients to detect
differences. The upper bound for the CI lay below the lower bound for the CIs estimated from Buse et al,48 DeFronzo et al,49 and Kendall et al 50 (Figure 3). The exenatide result from AWARD-1 appears to be inconsistent with the other trials. One potential explanation for the difference in effects across these trials is lack of blinding for the EBID arm in AWARD-1: placebo was given QW to match the Dula arms, but there was no dummy-placebo injection to match the BID dosing for exenatide.

In HARMONY-2, it was noted that the 50 mg Albi arm was not as effective as the 30 mg Albi arm at achieving the <7% A1C target (40.2% versus 49%). Note that the data for the HARMONY-2 study were taken from the clinical trials registry, as the study was not published in full at the time of this analysis. There may be unreported factors that would explain this result, but our analysis reflects the data as they were reported at the time of writing.

**Discussion**

In this NMA, we combined direct and indirect evidence from 29 RCTs involving 18,542 patients with T2D to estimate the relative efficacy among licensed GLP-1 RAs on the gold standard measure of diabetes control – A1C. We made several key observations: 1) GLP-1 RAs were superior to placebo in improving A1C, with moderate confidence in estimates; 2) relative efficacy was similar among longer-acting GLP-1 RAs, and absolute reduction in A1C at 6 months was consistent among Dula, EQW, and Lira and was estimated to be within the range 0.9%–1.4%; and 3) using NNT, we estimated that for every two patients treated with a GLP-1 RA, one will meet the <7% A1C target within 6 months.

Compared to direct evidence from head-to-head studies, evidence generated from this NMA allows for a more accurate assessment of GLP-1 RA relative efficacy on a class-wide level, which is especially important for population-health decision makers.21 Although direct evidence can provide health care decision makers with a crude sense of relative efficacy, such comparisons often lack details on the relative magnitude of treatment effects, are biased due to open-label trial design, and lack statistical power due to small sample size. The inconsistent results yielded from numerous direct comparisons also make it difficult for health care decision makers to reach definitive conclusions on a treatment decision.

While NMAs are useful in quantifying treatment effects from clinical studies, meta-analysis results can vary depending on the trials included in the network, the statistical methods applied, and the consideration of covariates to account for trial-design differences. Our network analysis was a comprehensive analysis of the A1C outcome across all US-licensed GLP-1 RAs. Our results are corroborated by other published meta-analyses of GLP-1 RAs, which showed that patients with T2D can expect to improve their A1C with GLP-1 RA therapy.14,16,19 However, our analysis completes the evidence for GLP-1 RAs, since other analyses either did not cover all currently available GLP-1 RAs in the US,13–17,19 limited study inclusion to placebo or specific active controls or combinations,14,16,17 or did not conduct an NMA to provide
comparisons of all GLP-1 RAs against one another.\textsuperscript{13,15–17} Furthermore, some studies did not use an established NMA method, such as the Bayesian method recommended by the NICE Decision Support Unit,\textsuperscript{21} and/or the NMA did not control for study-arm baseline $\text{A}_{1c}$ which varied across RCTs.\textsuperscript{18} Although baseline $\text{A}_{1c}$ was not found to be a statistically significant covariate in this study, other studies have shown that baseline $\text{A}_{1c}$ value could impact the magnitude of $\text{A}_{1c}$ reduction where higher reductions are associated with higher baseline $\text{A}_{1c}$ values.\textsuperscript{32}

The implications of consistent glycemic control help clinicians design individualized treatment plans. While new drug therapies target the multiple defects that together contribute to diabetes, the possibility of improved control through GLP-1 RAs may lead to improved patient outcomes beyond glycemic control. American Association of Clinical Endocrinologists guidelines recommend initiating treatment with metformin in patients with entry $\text{A}_{1c}>7.5\%$ plus a second agent, with preference given to treatments with low potential for hypoglycemia and weight-loss effects.\textsuperscript{67} Therefore, GLP-1 RAs are ranked hierarchically first before other options in this regard. In addition, increased $\text{A}_{1c}$ has been associated with microvascular and macrovascular complications, and lowering $\text{A}_{1c}$ to below or around $7\%$ has been shown to reduce microvascular and neuropathic complications of T1D and T2D.\textsuperscript{68,69} Most recently, this class has been shown to reduce major adverse cardiac events in large cardiovascular-outcome trials. Altogether, these findings, along with the known glycemic effects and beneficial secondary effects, may compel clinicians to use these agents in patients requiring robust and sustained control of their hyperglycemia, while addressing concerns of weight gain and hypoglycemia typically seen with traditional agents.

While important, the impact of a GLP-1 RA on $\text{A}_{1c}$ reduction is just one consideration when selecting the best treatment for a patient. Karagiannis et al\textsuperscript{16} provided a direct meta-analysis of weekly GLP-1 RAs that also covered other important outcomes, such as weight change and gastrointestinal and injection-site reactions. However, the analysis did not provide head-to-head or indirect comparisons, and there were insufficient results for EQW, which had a large clinical program design comparable with others in the class. A recent NMA by Sun et al\textsuperscript{18} focused on gastrointestinal adverse events, specifically nausea, vomiting, and diarrhea, and indicated that these effects are associated with GLP-1 RAs. More recently, Zaccardi et al\textsuperscript{19} performed an NMA of weekly GLP-1 RAs using the frequentist approach, and reported no differences between EQW and a maintenance dose of Dula (1.5 mg) for $\text{A}_{1c}$ or on all three metabolic outcomes (blood pressure, blood lipids, and C-reactive protein), and both treatments reduced $\text{A}_{1c}$ to a greater extent than albiglutide.

Managing blood glucose is fundamental to caring for people with T2D. With newer glucose-dependent agents, such as GLP-1 RAs, CER is an increasingly important tool, given the wide range of treatment options available in each class. It allows health care decision makers to evaluate the efficacy and safety of multiple treatment options simultaneously. Such methods as NMA have been well recognized as useful tools to evaluate the relative merits of treatments when direct head-to-head studies are not available.

**Limitations**

In this comparative-effectiveness analysis, the NMA method was used to integrate placebo- and active-controlled trial data to assess the relative efficacy of US-approved GLP-1 RAs. The NMA approach has an advantage in that it preserves randomized comparisons and gives each trial an appropriate weighting, while including data from both direct (head-to-head) studies and indirect studies (eg, via placebo). Inferences that are based on the direct evidence alone ignore a substantial part of the available clinical evidence. An NMA that includes both direct and indirect evidence provides a more comprehensive assessment of efficacy, and is less prone to study-selection bias. The value of NMA to CER is that it allows us to assess the magnitude of an intervention’s effect and its consistency across trials, as opposed to a “vote-counting” approach, which infers the presence, or not, of an effect based on the statistical significance of results in each study.\textsuperscript{36,70} Despite its strengths, this method is not without challenges.\textsuperscript{71,72} Therefore, our findings should be interpreted in light of the following limitations. One criticism of NMAs is that the methodology can lack transparency and results can be difficult to reproduce. Recent guidelines have helped standardize methods to improve confidence in NMAs.\textsuperscript{73} For our analysis, we followed the NICE Decision Support Unit recommendations that were developed in collaboration with leading academics for conducting and reporting NMAs, and used validated code that is available in the public domain.\textsuperscript{21}

Another limitation is that the NMAs rely on the assumption that data are consistent across trials. However, this is a problem associated with data synthesis in general and not just (network) meta-analysis. Problems with consistency may arise if the inclusion criteria are too broad, such that the trial populations are not comparable clinically. For example, treatment-naive patients may have a higher response to treatment compared with patients for whom one or more lines of
treatment have failed. There may also be undetected heterogeneity across trials that may arise from study bias, eg, poor quality, or small study bias where the trial results appear to be outliers. Some heterogeneity is to be expected, and some differences across clinical studies may reflect differences in real-world practice. There may be imbalances in the distribution of unobserved or unmeasured effect modifiers that have the potential to confound the comparative estimates among GLP-1 RAs. Also, we did not consider whether differences between safety outcomes across GLP-1 RAs impact on treatment efficacy, though this has been considered elsewhere.18

Finally, our NMA included non-GLP-1 RA drugs (eg, metformin, sulfonylureas, insulin) as control arms or additional arms from the GLP-1 RA trials. These arms are required to connect the network of evidence across the GLP-1 RA trials that are the focus of this analysis. While our analysis suggests lower relative efficacy of non-GLP-1 RAs versus GLP-1 RAs, the analysis does not include all available evidence for the non-GLP-1 RA drugs.

Conclusion
This is a comprehensive assessment of the comparative effectiveness of US-licensed GLP-1 RAs in terms of A1c. GLP-1 RAs are superior to placebo in improving glycemic control, with a consistent absolute reduction in A1c at 6 months, ranging from 0.9% to 1.4%, among Dula, EQW, and Lira. In terms of NNT, we estimate that for every two patients treated with a GLP-1 RA, one will meet the <7% A1c target within 6 months of commencing GLP-1 RAs. These GLP-1 RAs in particular should thus be considered a viable addition to oral antidiabetes therapy in the appropriate patient.

Disclosure
MEO received funding from AstraZeneca US, and HN, JL, and SAT are paid employees of AstraZeneca US. The authors report no other conflicts of interest in this work.

References


