U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research
Office of Translational Sciences
Office of Biostatistics

STATISTICAL REVIEW AND EVALUATION

CLINICAL STUDIES

NDA/BLA #:	NDAs 204629, 206111, 208658
Supplement #:	S-42, S-38, S-26
Drug Name:	Jardiance (empagliflozin), Synjardy (empagliflozin + metformin), Synjardy XR (empagliflozin and metformin extended release)
Indication(s):	To improve glycemic control in pediatric patients (10 to 17 years) with Type 2 diabetes mellitus
Applicant:	Boehringer Ingelheim Pharmaceuticals, Inc (BIPI)
Date(s):	Receipt dates: December 20, 2022 (for Jardiance), January 31, 2023 (for Synjardy), April 25, 2023 (for Synjardy XR) Review due date: May 26, 2023 6-Month PDUFA Goal date: June 20, 2023
Review Priority:	Priority
Biometrics Division:	DB II
Statistical Reviewer:	Wenda Tu (Primary statistical reviewer) Satyajit Ghosh (Pediatric statistical reviewer for Bayesian analysis)
Concurring Reviewers:	Yoonhee Kim (Diabetes statistics team TL) James Travis (Pediatric statistics team TL)
Medical Division:	DDLO
Clinical Team:	Kim Shimy/Michelle Carey (TL)
Project Manager:	Michael Oyewole

Keywords:

Type 2 diabetes, pediatric study, SGLT-2 inhibitor, master protocol, shared control, analysis of covariance, placebo washout imputation, shrinkage analysis, Bayesian hierarchical modelling

Table of Contents

LIST OF TABLES

LIST OF FIGURES

1 EXECUTIVE SUMMARY

On December 20, 2022, the Sponsor, BIPI, submitted NDA 204629 S-042 for Jardiance (empagliflozin, or empa) and NDA 201280 S-027 for Tradjenta (linagliptin, or lina), in support of product label updates with respect to the pediatric indication. The label updates of both products were based on a single Phase 3 pediatric trial titled "DIabetes Study of LiNAgliptin and eMpagliflozin in children and adOlescents" (DINAMO). The study was conducted to satisfy the pediatric PMR-3300-1, which applies to all drug products containing empa and lina.

On January 31, 2023, BIPI submitted NDA 206111 S-038 for Synjardy (the fixed dose combination product [FDC] of empa and metformin, or empa + met), and NDA 201281 S-035 for Jentadueto (the FDC of lina and metformin, or lina + met). On April 25, 2023, BIPI submitted NDA 208658 S-026 for Synjardy XR (empa + met extended release [XR]), and NDA 208026 S-024 for Jentadueto XR (lina + met XR). All four NDA supplements referred to the study DINAMO for product label updates regarding the pediatric indication. To facilitate the review, the Agency decided to combine the internal review timelines of all the aforementioned NDA supplements concerning empa and lina. This statistical review focuses on empa, its FDC with metformin, and its FDC with metformin XR, under NDA204629, NDA206111, and NDA 208658, respectively. Refer to a separate review for lina, its FDC with metformin, and its FDC with metformin XR under NDA201280, NDA 201281, and NDA 208026.

The three drug products containing empa are currently indicated for treatment of adult patients with type 2 diabetes mellitus (T2DM) as adjuncts to diet and exercise. In the current submissions, the Sponsor proposed to expand the T2DM indication to the pediatric population aged 10 to 17 years. The proposed label update was based on the analysis result from the Study DINAMO, which demonstrated a statistically significant treatment effect for empa vs placebo with respect to the primary endpoint HbA1c change from baseline at Week 26; i.e., the placebo-adjusted HbA1c change from baseline was -0.84% with a 95% confidence interval (-1.50, -0.19).

1.1 Brief overview of Clinical Study

The Study DINAMO was a multi-center, randomized, parallel-group, placebo-controlled study intended to evaluate the efficacy and safety of lina 5 mg and an empa dosing regimen vs. placebo after 26 weeks of treatment in children and adolescents with T2DM. It consisted of 1-week Screening Period, a 2-week Run-in Period, a 26-week Main Treatment Period, a 26-week Extended Treatment Period, and a 3-week safety Follow-up Period. At Week 1 of the Main Treatment Period, a total of 158 subjects were randomized in a 1:1:1 ratio to one of the three treatment arms: empa 10 mg, lina 5mg, or placebo. At Week 14, non-responders to empa 10 mg[1] underwent a second randomization to either empa 10 mg or empa 25 mg in a 1:1 ratio. The primary endpoint HbA1c change from baseline was assessed at Week 26 of the Main Treatment Period.

[1] Non-responders to empa 10 mg refer to subjects who were randomized to empa 10 mg at Week 1 but failed to achieve HbA1c < 7 % assessed at Week 12.

1.2 Major Statistical Issues

No major statistical issues have been identified in this review. The overall missing rate was 9.6% for empa, and 5.7% for placebo. Missing endpoints were multiply imputed based on placebo washout. For primary efficacy analyses, the applicant applied an ANCOVA adjusted for treatment (placebo, empa or lina), baseline HbA1c and age stratum at baseline (< 15 years vs 15 to <18 years).

Minor review issues were identified as follows. Firstly, the study was designed under a master protocol with two active drugs (empa and lina) vs a shared placebo arm. Since the primary hypothesis tests concerned two independent tests of distinct drugs vs. placebo, there is no need for multiplicity adjustment (Section 3.2.2). The Sponsor, however, used the Hochberg procedure to control the overall Type 1 error rate for comparing two active drugs against a shared placebo arm. Secondly, after the second randomization at Week 14, no dose-response relation was observed among the non-responders to empa 10 mg (Section 3.2.4). The submission package also included the results of the supplementary Bayesian borrowing analyses, which the Sponsor has agreed to perform in order to address the Agency's concern about the study sample size. Details about the Bayesian review can be found in the appendix (Section 6).

1.3 Collective Evidence

The primary efficacy results are summarized in Table 1. Additionally, results from sensitivity analyses demonstrated robustness of the primary efficacy results to untestable assumptions on missing data (Section 3.2.4). Subgroup analyses on the primary efficacy endpoint found consistent treatment effect of empa in subgroup levels based on age, sex, race, and region (Section 4.1), as well as background medications (Section 4.2). An elevated risk of hypoglycemia was found in subjects treated with empa compared to those treated with placebo (Section 3.3).

Table 1: Primary Efficacy Result on HbA1c Change from Baseline at Week 26

	Empa pooled N=52	**Placebo N=53**
Baseline, mean (SD)	8.00 (1.29)	8.05 (1.23)
Missing primary endpoint, n (%)	5 (9.6)	3 (5.7)
Change from baseline, LSMean[1] (SE)	-0.17 (0.24)	0.68 (0.23)
Difference from Placebo, LSMean[1] (CI)	-0.84 (-1.50, -0.19)	
Two-sided p-value (unadjusted)	0.01	

Abbreviations: CI = confidence interval, SD = standard deviation, SE = standard error.
[1] The LSMean estimate is based on an ANCOVA model adjusted for baseline HbA1c, baseline age stratum (< 15 years vs 15 to <18 years), and treatment after imputing missing data using placebo washout method
Source Clinical Study Report (CSR) Table 15.2.1.1 1 (Page 312)

1.4 Conclusion and Recommendations

Statistical analyses based on the clinical data from the Phase 3 pediatric study DINAMO have demonstrated robust evidence to support the effectiveness of empagliflozin regarding glycemic control among pediatrics (11 to < 18 years) with T2DM. The statistical review team recommend approval of the proposed label updates for Jardiance, Synjardy and Synjardy XR.

2 INTRODUCTION

2.1 Overview

Empagliflozin (Jardiance®), a sodium-glucose co-transporter 2 (SGLT2) inhibitors, and its FDC with metformin (Synjardy® and Synjardy® XR) were approved by the FDA in 2014 and 2015, respectively, both as adjuncts to diet and exercise to improve glycemic control among adults with T2DM. In the current NDA supplements, the applicant proposed to expand the indication of both Jardiance and Synjardy to pediatric patients (aged 10 to 17 years) with T2DM. The proposed label updates were based on the analysis results from the Phase 3 study DINAMO conducted among pediatric patients with T2DM aged 11 to 17 years. The study started on April 26, 2018 and completed on June 27, 2022. Database lock occurred on August 10, 2022. An overview of the study is presented in Table 2.

Table 2: Overview of Study DINAMO

Trial ID	Design*	Treatment (Sample size)	Endpoint/Analysis
1218.91	MC, R, DB, PG, PC (3-week screening & run-in period + 52-week treatment period + 3-week follow-up period)	Empagliflozin 10 mg and 25 mg† (empa pooled) (N = 52) Linagliptin 5mg (lina) (N = 52) Placebo (pbo) (N = 53)	Primary: Change in HbA1c from baseline at Week 26 Key Secondary: None The primary endpoint was analyzed with an ANCOVA adjusted for treatment, baseline HbA1c, and baseline age category (< 15 years vs 15 to < 18 years). The analysis was based on the mITT population††, with missing data multiply imputed using the washout method.

* MC: multi-center, R: randomized, DB: double-blind, PG: parallel group, PC: placebo controlled
† Subjects from the empa arm received empa 10mg at the beginning of the Treatment Period. At Week 12, an evaluation of HbA1c was performed for these subjects. Subjects achieved HbA1c < 7.0% continued with empa 10mg, whereas those who failed the A1c target were further randomized to empa 10mg or empa 25mg at Week 14. The primary efficacy analysis was conducted based on the empa (10 mg and 25 mg pooled) vs placebo .
†† The mITT population was defined as all subjects who were randomized and received treatment

2.2 Data Sources

The Electronic Document Room (EDR) location for the Jardiance submission package is \\CDSESUB1\evsprod\NDA204629\1321\. Datasets for the study DINAMO (both in ADAM format and SDTM format) and the programming codes for the efficacy analyses can be found

under the subdirectory: m5\datasets\1218-0091. The EDR location for the Synjardy package is \\CDSESUB1\evsprod\NDA206111\0468.

On March 3, 2023, an IR was sent to the Sponsor requesting additional subgroup analyses based on background medication. The Sponsor's response can be found at \\CDSESUB1\evsprod\NDA204629\1494

On March 14, 2023, an IR was sent to the Sponsor requesting efficacy analyses on FPG change from baseline at Week 26 and BMI Z-score change from baseline. The Sponsor's response can be found at \\CDSESUB1\evsprod\NDA204629\1509.

On April 18, 2023, an IR was sent to the Sponsor requesting model-based analyses on hypoglycemia event counts. The Sponsor's response can be found at \\CDSESUB1\evsprod\NDA204629\1546.

3 STATISTICAL EVALUATION

3.1 Data and Analysis Quality

No issues have been identified with respect to data and analysis quality.

3.2 Evaluation of Efficacy

3.2.1 Study Design and Endpoints

The study DINAMO was a multi-center, randomized, parallel-group, placebo-controlled study intended to evaluate the efficacy and safety of lina 5 mg and an empa dosing regimen vs. placebo after 26 weeks of treatment in children and adolescents with T2DM. As demonstrated in Figure 1, the study consisted of a one-week Screening Period, a two-week Run-in Period, a 26-week Main Treatment Period, a 26-week Extended Treatment Period, and a three-week safety Follow-up Period. At Week 1 of the Main Treatment Period, a total of 158 subjects were randomized in a 1:1:1 ratio to one of the three treatment arms: empa 10 mg, lina 5mg, or placebo. The randomization was stratified by age (< 15 years vs 15 to < 18 years).

At Week 12, subjects on empa 10 mg were assessed for their HbA1c levels. Those who failed to achieve HbA1c < 7% (i.e., non-responders to empa 10mg) underwent a second randomization at Week 14, during which subjects were randomized to either empa 10 mg or empa 25 mg in a 1:1 ratio. The primary endpoint HbA1c change from baseline was assessed at the end of the Main Treatment Period (i.e., Week 26). At the beginning of the Extended Treatment Period followed (Week 26 – 52), subjects previously on placebo were randomized to lina 5mg, empa 10mg or empa 25mg in a 1:1:1 ratio, whereas subjects previously on active treatment continued with their treatment. A three-week safety assessment followed at the conclusion of the Extended Treatment Period.

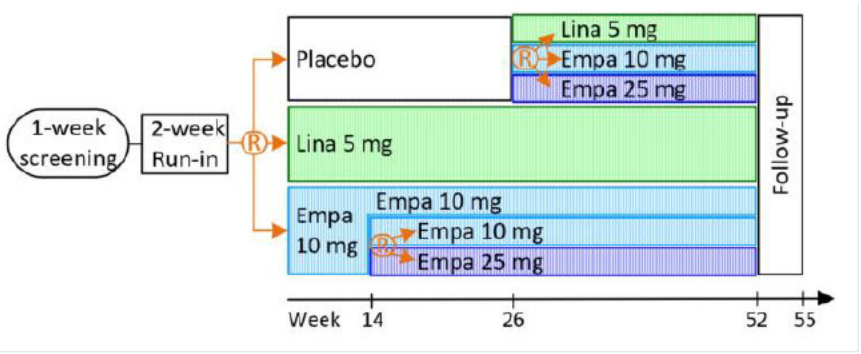


Figure 1: Trial Design for DINAMO
Source: Figure 9:1, CSR

The primary objective of the study was to demonstrate both superiority of empa (10 mg and 25 mg pooled) to placebo, and superiority of lina 5 mg to placebo, as assessed by the primary endpoint: HbA1c (%) change from baseline at Week 26. The study did not specify any key secondary endpoint.

Sample Size

The determination of the study sample size, as initially specified in the SAP, is as follows. Assuming a -0.55% treatment effect difference between the active treatment group (lina or empa) and the placebo group and a 0.9% standard deviation (SD), a sample size of 50 subjects per initial randomized treatment arm (150 subjects in total) would provide 85% power at a two-sided 0.05 level.

In the Information Request (IR) Letter issued on November 16, 2021, the Agency expressed concerns that the study might be undersized due to the consideration that the observed SD might be greater than the assumed SD of 0.9%. This concern stemmed from recently completed pediatric T2DM trials in which the SDs were generally found larger than adult T2DM trials. In response to this IR, the Sponsor conducted a blinded interim check of the SD. At the time, 157 subjects had started treatment and 141 subjects were included in the SD calculation. The SD observed from the interim check was 1.65%, which confirmed the Agency's concern. The Sponsor, nonetheless, refused to increase the sample size by arguing that the assumed effect size of 0.55% was too conservative for empa. Based on the estimated treatment effect for dapagliflozin from the pediatric T2DM trial NCT072725593, a more likely treatment effect for empa (after placebo adjustment) should be -0.87%. With this new assumption for the treatment effect and an observed SD of 1.65%, a sample size of 52 per arm would provide 75% power at a two-sided 0.05 level. The Agency agreed with the Sponsor's decision of no sample size increase, but asked the Sponsor to perform a supplemental Bayesian borrowing analysis as additional supportive evidence to address the sample size concern. The requested analyses were submitted in NDA 204629, S-42, and was reviewed by Dr. Satyajit Ghosh from the Pediatrics and Maternal Health Team at DB II (Section 6).

In reality, 52 subjects on empa pooled and 53 subjects on placebo were randomized and treated in the study. The pooled SD for the empa and the placebo groups was 1.71%, and the estimated

treatment effect was -0.84% for empa after placebo adjustment. A statistically significant treatment effect was found for empa at a two-sided 0.05 level.

Primary Endpoint

- Change from baseline in HbA1c (%) at Week 26

Secondary Endpoints

- Change from baseline in fasting plasma glucose at Week 26
- Change from baseline in body weight at Week 26
- Change from baseline in systolic blood pressure at Week 26
- Change from baseline in diastolic blood pressure at Week 26
- Incidence of HbA1c < 6.5% at Week 26
- Incidence of HbA1c < 7.0% at Week 26

3.2.2 Statistical Methodologies

The Sponsor did not pre-specify an estimand framework for statistical analyses in SAP. The key components of an estimand are summarized as follows based on the pre-specified statistical approaches used for the primary efficacy analysis.

Population & Analysis Set

The target population was the modified ITT (mITT) population, defined as all randomized and treated subjects who had baseline HbA1c measurements, regardless of treatment adherence or rescue medication.

Handling of Missing Data

Multiple imputation based on placebo washout was applied. Specifically, missing data from the placebo arm were imputed with a sequential linear regression constructed based on observed HbA1c values from the placebo group, measured at baseline, Weeks 4, 12 and 26. Missing data from the treatment arm were imputed with a sequential linear regression constructed based on the observed HbA1c values from the placebo group, measured at baseline and Week 26. 1000 imputed dataset were created, and Rubin's Rule was used to combine the inference results.

Multiplicity Adjustment

The two primary hypotheses concern comparisons of empa pooled against placebo and lina 5 mg against placebo with respect to the primary endpoint: HbA1c change from baseline at Week 26. To control the overall Type I error rate at a two-sided 0.05 level, the Sponsor applied the Hochberg procedure for simultaneous testing of the two primary hypotheses.

After having obtained statistically significant results for both primary hypotheses, two secondary hypotheses that compare different empa regimen groups against placebo were tested on the primary endpoint at a two-sided 0.05 level. The tests were conducted sequentially in the following order:

1. TITR 25 + responders (Non-responders to empa 10 mg who were titrated to 25 mg at Week 14 + Responders to empa 10 mg) vs. placebo
2. TITR10 + responder (Non-responders to empa 10 mg who continued with empa 10 mg at Week 14 + Responders to empa 10 mg) vs. placebo,

Reviewer's note:
Based on the study result, the primary hypothesis test on lina 5 mg vs placebo failed; hence no formal testing for the secondary hypothesis family was performed. However, according to the recently published FDA Guidance on master protocol for oncology product development[2], multiplicity adjustment is considered unnecessary for multiple comparisons of different investigational drugs to the comparator group in an umbrella trial setting. This suggested that the two primary hypothesis tests in this study can be conducted independently, each at a two-sided 0.05 level. Further, the failure of the primary hypothesis test on lina vs placebo should not preclude a formal testing of the secondary hypothesis family, as both secondary hypotheses concern comparisons of empa vs. placebo. Given this consideration, a more efficient testing structure for this study should involve two testing sequences. The first sequence consists of a single test for lina 5 mg vs placebo, whereas the second sequence involves a sequential testing procedure started with empa pooled vs. placebo, and followed by the two secondary hypothesis tests on the empa subgroups vs. placebo.

Despite a more efficient use of alpha, the two-sequence testing structure would still yield the same conclusions as the current testing structure, however. This is because the first hypothesis test in the secondary hypothesis family failed (i.e., LSMean [95% CI] for placebo-adjusted A1c change from baseline at Week 26 among TITR 25 + responders: -0.52 [-1.31, 0.27]). Therefore, the empa testing sequence would stop at this test and claim success only on the primary hypothesis test of empa pooled vs placebo.

Primary Efficacy Analyses
The primary hypothesis test was performed based on an ANCOVA, with HbA1c change from baseline at Week 26 as the response variable, and treatment, baseline HbA1c, and baseline age category (< 15 years vs 15 to < 18 years) as covariates.

The secondary hypothesis family intended to explore the question of whether non-responders to empa 10mg would benefit from a dose up-titration to empa 25mg. Each hypothesis test from the secondary hypothesis family was performed based on the same ANCOVA as for the primary hypothesis test, but with the application of the inverse probability weighting (IPW) technique. To explain how IPW works, consider comparing empa 10mg (without dose up-titration) vs. placebo as an example. At the beginning of the study, a weight variable ω was created for each subject. All subjects started with $\omega = 1$. At Week 14, the empa non-responders who were up-titrated to empa 25mg (TITR 25) would have their weights transferred to the empa non-responders who were randomized to continue with empa 10mg (TITR 10) (*i.e.*, the TITR 10 group had $\omega = 2$,

[2] FDA Guidance for Industry: Master Protocols: Efficient Clinical Trial Design Strategies to Expedite Development of Oncology Drugs and Biologics: https://www fda.gov/media/120721/download

and the TITR 25 group had $\omega = 0$). This way, the TITR 25 group were represented by the TITR 10 group. All other subjects not involved in the second randomization had $\omega = 1$. The diagonal matrix **W** was created accordingly and used in the ANCOVA model as the weight matrix[3]:

$$\boldsymbol{W} = \begin{bmatrix} \boldsymbol{I}_n & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{rn} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & 2 * \boldsymbol{I}_{\frac{(1-r)n}{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} * \boldsymbol{I}_{\frac{(1-r)n}{2}} \end{bmatrix}$$

In the matrix **W**, **I** is an identity matrix, with its dimension specified by the subscript. n is the sample size for each treatment arm, r is the proportion of responders in empagliflozin arm. $\mathbf{I}_n$ indicates $\omega = 1$ for all subjects in the placebo arm; $\boldsymbol{I}_{rn}$ indicates $\omega = 1$ for empa responders; $2 * \mathbf{I}_{\frac{(1-r)n}{2}}$ indicates $\omega = 2$ for empa non-responders randomized to remain on empa 10mg; and $0 * \mathbf{I}_{\frac{(1-r)n}{2}}$ indicates $\omega = 1$for empa non-responders up-titrated to empa 25mg.

A similar weighting scheme was applied for the comparison of empa up-titration to 25mg vs. placebo, where the transfer of weight was from TITR 10 to TITR 25. Since the two hypothesis tests from the secondary hypothesis family share the same subset of empa responders, the comparisons of TITR 10 and TITR 25 to placebo are highly correlated. This allows a reduction in sample size from the design that initially randomizes subjects to two different empa doses.

Sensitivity Analysis

In the Sponsor's submission package, a mixed model for repeated measure (MMRM) based on the mITT population was used as a sensitivity analysis for the confirmatory tests of the primary hypothesis family. This is considered insufficient from a regulatory perspective, as an MMRM assumes data are missing at random, which is an unlikely scenario for many missing cases in clinical trials. In this review, to study the impact of missing data on the primary analysis result, the primary endpoint was modeled with the same ANCOVA as the Sponsor's, while missing primary endpoints were multiply imputed based on the return-to-baseline approach.

3.2.3 Patient Disposition, Demographic and Baseline Characteristics

A summary of subject disposition is presented in Table 3. All randomized subjects received at least one dose of the study drug. No notable difference was observed between the empa and placebo arms with respect to the study disposition. The placebo arm has a slightly higher treatment discontinuation rate than the empa arm (20.8% vs. 15.4%), mainly driven by a higher rate of treatment withdrawal. Five subjects from the empa arm and three subjects from the

[3] This was implemented in the SAS procedure PROC MIXED, with the WEIGHT statement specified as the weight matrix.

placebo arm missed their primary endpoint assessments. Due to a limited sample size of the retrieved dropouts, missing data cannot be imputed based on the retrieved dropout group.

Table 3: Subject Disposition

	Empa pooled (%) N = 52	**Placebo (%) N = 53**	**Total (%) N = 105**
Treated	52 (100)	53 (100)	105 (100)
Study Discontinuation (up to Week 52)	7 (13.5)	6 (11.3)	13 (12.4)
Lost to follow-up	1 (1.9)	2 (3.8)	3 (2.9)
withdrawal by subject	5 (9.6)	4 (7.5)	9 (8.6)
Other	1 (1.9)	0	1 (1)
Treatment Discontinuation (up to Week 52)	8 (15.4)	11 (20.8)	19 (18.1)
Adverse Event	0	2 (3.8)	2 (1.9)
Lost to follow-up	0	1 (1.9)	1 (1)
Withdrawal by subjects	4 (7.7)	7 (13.2)	11 (10.5)
Other	4 (7.7)	1 (1.9)	5 (4.8)
Primary Endpoint Missing (up to Week 26)	5 (9.6)	3 (5.7)	8 (7.6)
Retrieved Dropout (up to Week 26)	3 (5.8)	2 (3.8)	5 (4.8)

Source Table 10 2, CSR, and reviewer's analysis; adsl.xpt, adhba1c.xpt

At Week 14, 47 subjects initially randomized to the empa 10mg were still on treatment. 24 (51%) of them were non-responders, and underwent a second randomization to either empa 25mg (N = 13), or empa 10 mg (N = 11).

A summary of patient demographics and baseline characteristics is presented in Table 4. Based on the summary, demographics and baseline characteristics are well balanced between the empa and placebo groups.

Table 4: Demographics and Baseline Characteristics

	Empa Pooled N=52	Placebo N=53	Total N=105
Sex , n (%)			
Female	33 (63.5)	34 (64.2)	67 (63.8)
Male	19 (36.5)	19 (35.8)	38 (36.2)
Age , years			
Mean (SD)	14.4 (1.94)	14.6 (1.76)	14.5 (1.85)
Median	15.0	14.0	14.0
Min, Max	10.0, 17.0	11.0, 17.0	10.0, 17.0
Age Category , n (%)			
<15	25 (48.1)	26 (49.1)	51 (48.6)
≥ 15 to <18	27 (51.9)	27 (50.9)	54 (51.4)
Race , n(%)			
American Indian/Alaska Native	4 (7.7)	1 (1.9)	5 (4.8)
Asian	2 (3.8)	3 (5.7)	5 (4.8)
Black/African American	19 (36.5)	17 (32.1)	36 (34.3)
Native Hawaiian/Pacific Islander	0	1 (1.9)	1 (1.0)
White	23 (44.2)	29 (54.7)	52 (49.5)
Multiple	4 (7.7)	1 (1.9)	5 (4.8)
Missing	0	1 (1.9)	1 (1.0)
Region, n (%)			
Asia	1 (1.9)	1 (1.9)	2 (1.9)
Europe	6 (11.5)	7 (13.2)	13 (12.4)
North America	36 (69.2)	34 (64.2)	70 (66.7)
South America	9 (17.3)	11 (20.8)	20 (19.0)
Baseline BMI Z-score, n (%)			
≥ -2 to 1 (Normal)	1 (1.9)	2 (3.8)	3 (2.9)
>1 to 2 (Overweight)	4 (7.7)	7 (13.2)	11 (10.5)
>2 (Obese)	47 (90.4)	44 (83.0)	91 (86.7)
Baseline background medication, n (%)			
Insulin Only	3 (5.8)	2 (3.8)	5 (4.8)
Metformin and Insulin	22 (42.3)	19 (35.8)	41 (39.0)
Metformin Only	26 (50.0)	28 (52.8)	54 (51.4)
None	1 (1.9)	4 (7.5)	5 (4.8)
Baseline HbA1c, %			
Mean (SD)	8.0 (1.29)	8.1 (1.23)	8.0 (1.25)
Median	7.9	7.6	7.9
Min, Max	6.2, 10.6	6.0, 10.7	6.0, 10.7
Baseline A1c category, n (%)			
<8.0	28 (53.8)	29 (54.7)	57 (54.3)
≥ 8.0	24 (46.2)	24 (45.3)	48 (45.7)

Source Statistical Reviewer Analysis; adsl.xpt

3.2.4 Results and Conclusions

As demonstrated in Table 5, the LSMean difference (95% CI) in HbA1c change from baseline at Week 26 is -0.84 (-1.50, -0.19) for empa pooled vs. placebo, with a two-sided p-value 0.01. The study has successfully demonstrated superiority of empa to placebo with respect to glycemic control.

Table 5: HbA1c Change from Baseline at Week 26, Primary Hypothesis

	Empa pooled N=52	**Placebo N=53**
Baseline, mean (SD)	8.00 (1.29)	8.05 (1.23)
Missing primary endpoint, n (%)	5 (9.6)	3 (5.7)
Change from baseline, LSMean[1] (SE)	-0.17 (0.24)	0.68 (0.23)
Difference from Placebo, LSMean[1] (CI)	-0.84 (-1.50, -0.19)	
Two-sided p-value (unadjusted)	0.01	

Abbreviations: CI = confidence interval, SD = standard deviation, SE = standard error.
[1] The LSMean estimate is based on an ANCOVA model adjusted for baseline HbA1c, baseline age stratum (< 15 years vs 15 to <18 years), and treatment. Missing data was multiply imputed based on the method of placebo washout. Inference results were combined with Rubin's Rule.
Source Table 15.2.1.1 1, CSR

For sensitivity analysis, missing primary endpoint was multiply imputed based on the return-to-baseline approach. The same ANCOVA model as the primary efficacy analysis was fitted on 500 imputed datasets, and Rubin's Rule was applied to combine the inference results. As shown in Table 6, the placebo-adjusted treatment effect was -0.90 with a 95% confidence interval (-1.53, -0.27) and a two-sided p-value of 0.01. This has confirmed the conclusion based on the primary analysis.

Reviewer's Note
In diabetes trials, the return-to-baseline method is generally considered a more conservative imputation method than the placebo washout method, as the former assumes that zero treatment effect is retained for subjects who missed primary endpoint, whereas the latter assumes that subjects who discontinue the active treatment can still benefit from the standard of care as administered in the placebo arm. In this study, however, subjects from the placebo arm tend to have worse-than-baseline glycemic level at Week 26. Hence, the estimated treatment effect based on the placebo washout method appears slightly more conservative than the return-to-baseline method. Regardless, the estimated treatment effects based on these two methods were highly similar and were statistically significant, which shows that the primary efficacy result is robust to missing data assumptions.

Table 6: HbA1c Change from Baseline at Week 26, Sensitivity Analysis

	Empa pooled N=52	**Placebo N=53**
Baseline, mean (SD)	8.00 (1.29)	8.05 (1.23)
Change from baseline, LSMean[1] (SE)	-0.25 (0.23)	0.66 (0.22)
Difference from Placebo, LSMean[1] (CI)	-0.90 (-1.53, -0.27)	
Two-sided p-value (unadjusted)	0.01	

Abbreviations: CI = confidence interval, SD = standard deviation, SE = standard error.
[1] The LSMean estimate is based on an ANCOVA model adjusted for baseline HbA1c, baseline age stratum (< 15 years vs 15 to <18 years), and treatment. Missing data was multiply imputed based on the method of the method of return to baseline. Inference results were combined with Rubin's Rule.
Source Reviewer's Analysis; ada1c.xpt, adsl.xpt

The analysis results for the secondary hypothesis family were presented in Tables 7 and 8. As the primary hypothesis test on lina vs placebo failed (refer to statistical review under NDAs 201280 and 201281), these analyses are considered exploratory. The placebo-adjusted treatment effect (95% CI) was -0.52 (-1.31, 0.27) for the non-responders titrated to the empa 25 mg, and -1.18 (-1.90, -0.45) for the non-responders continued with the empa 10 mg.[4] A reversed dose response was observed for empa 25 mg and empa 10 mg in this second randomization regimen.

Table 7: HbA1c (%) Change from Baseline at Week 26, TITR25 vs. Placebo

	TITR 25 + Responders N = 41	**Placebo N = 53**
Baseline, mean (SD)	7.80 (1.26)	8.05 (1.23)
Change from baseline, LSMean[1] (95% CI)	0.14 (-0.42, 0.71)	0.66 (0.12, 1.21)
Difference from Placebo, LSMean[1] (95% CI)		-0.52 (-1.31, 0.27)
Two-sided p-value (unadjusted)		0.19

Abbreviations: CI = confidence interval, SD = standard deviation, SE = standard error.
[1] The LSMean estimate is based on an ANCOVA model with the application of inverse probability weighting, adjusted for baseline HbA1c, baseline age stratum (< 15 years vs 15 to <18 years), and treatment. Missing data was multiply imputed based on the method of the method of return to baseline. Inference results were combined with Rubin's Rule.
Source Table 15.2.1.2, CSR

Table 8: HbA1c (%) Change from Baseline at Week 26, TITR10 vs. Placebo

	TITR 10 + Responders N=39	**Placebo N = 53**
Baseline, mean (SD)	7.92 (1.36)	8.05 (1.23)
Change from baseline, LSMean[1] (95% CI)	-0.49 (0.27)	0.68 (0.19, 1.17)
Difference from Placebo, LSMean[1] (95% CI)		-1.18 (-1.90, -0.45)
Two-sided p-value (unadjusted)		0.002

Abbreviations: CI = confidence interval, SD = standard deviation, SE = standard error.
[1] The LSMean estimate is based on an ANCOVA model with the application of inverse probability weighting, adjusted for baseline HbA1c, baseline age stratum (< 15 years vs 15 to <18 years), and treatment. Missing data was multiply imputed based on the method of the method of return to baseline. Inference results were combined with Rubin's Rule.
Source Table 15.2.1.2, CSR

To further investigate this reversed trend in dose response, an ANCOVA without the application of IPW was applied to compare the treatment effect of TITR 10 vs TITR 25. The ANCOVA was based on data from empa non-responders only, with the response variable HbA1c change from Week 12[5] at Week 26. As presented in Table 9, a similar reversed dose response was observed. The CIs for TITR10 and TITR25 were highly overlapped, however, suggesting that the reversed dose response may be due to chance.

[4] The placebo-adjusted HbA1c change from baseline (95% CI) based on ANCOVA without IPW was -0.62 (-1.40, 0.16) for TITR25, and -1.08 (-1.93, -0.33) for TITR10.
[5] Since no HbA1c assessment was scheduled at Week 14, HbA1c at Week 12 were used as substitutes for Week 14.

Table 9: HbA1c Change from Week 12 at Week 26, Empa Non-Responders Only

	TITR 10 N = 11	TITR 25 N=13
Change from Week 12, LSMean[1] (95% CI)	-0.14 (-0.72, 0.45)	0.41 (-0.16, 0.98)

[1] The LSMean estimate is based on an ANCOVA model, adjusted for baseline HbA1c, baseline age stratum (< 15 years vs 15 to <18 years), and treatment. Missing data was multiply imputed based on the method of the method of return to baseline. Inference results were combined with Rubin's Rule.

Source reviewer's analysis; ada1c,xpt, adsl.xpt

It is worth noting that the non-responder group, by definition, consisted of subjects who failed to meet the glycemic target when treated with empa 10mg. The mean HbA1c change from baseline at Week 12 were -0.30% for the non-responders, as opposed to -1.00% for the responders. Hence, while empa 10mg appears generally effective among the full study population, its efficacy seems limited among the non-responder group. As displayed in Table 10, when compared to the responder group, the non-responder group on average had a higher HbA1c at baseline, a higher percentage of subjects with HbA1c $\geq$ 8.5% at baseline, and a higher percentage of subjects on a more aggressive background treatment regimen (i.e., metformin + insulin). All these facts suggested that even prior to randomization, the non-responders tend to have more advanced T2DM than the responders, which may explain the lack of responsiveness to the empa treatment observed in the empa non-responder group.

Table 10: Empa Non-Responders vs Responders

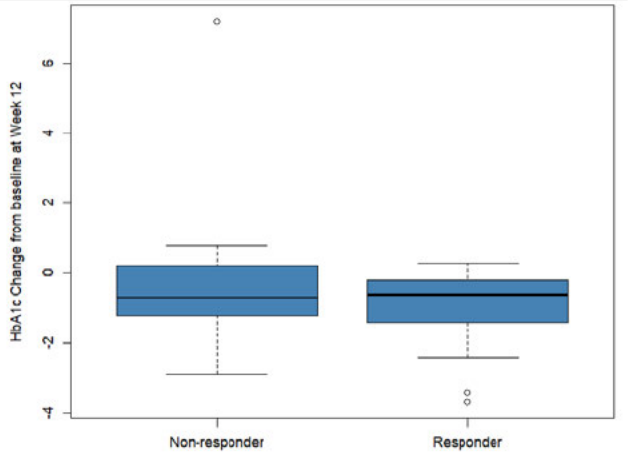
	Empa non-responders (N=24)	Empa responders (N=25)*
Baseline HbA1c (%), mean (SD)	8.5 (1.11)	7.2 (1.09)
HbA1c $\geq$8.5%	45.8%	12.0%
On Background metformin and insulin	58.3%	20.0%
A1c Change from Baseline at Week 12, mean (SD)	-0.30 (1.86) †	-1.00 (1.04)

* The original sample size is N = 28. 3 subjects in responder group missed Week 12 measurements, and hence were not counted for results presented in this table.

† The large SD was caused by the presence of outliers on the high end. Figure 2 displays the distribution of A1c change from baseline at Week 12.

Source reviewer's analysis; ada1c,xpt, adsl.xpt

Figure 2: Boxplots for HbA1c Change from Baseline at Week 12, Non-Responder vs Responder



Source reviewer's analysis; ada1c,xpt, adsl.xpt

Moreover, a dose response trend for empa was later observed during the Extended Treatment Period among subjects who were on placebo from Week 0 to Week 26, and were re-randomized to empa 10mg and empa 25mg at Week 26 (Table 11). Because a dose response was observed among the re-randomized subjects from the placebo group, the lack of dose response at Week 26 after the second randomization is likely a peculiar case for the empa non-responder group only.

Table 11: Descriptive Statistics of HbA1c (%) Change from Week 26 to Week 52, Among Subjects Previously on Placebo

	N	Week 26, mean (SD)	Change from Week 26, mean (SD)		
			Week 30	Week 42	Week 52
Empa 10 mg	15	8.43 (2.38)	-0.48 (0.65)	0.12 (1.41)	-0.35 (1.50)
Empa 25 mg	16	8.53 (2.37)	-0.60 (0.85)	-0.57 (1.05)	-0.53 (1.13)

Source Table 11 4, CSR

Besides the primary endpoint, analysis results for the secondary endpoints FPG change from baseline at Week 26, and BMI Z-score change from baseline at Week 26 were presented in Tables 12 and 13[6]. A nominally significant difference was found between the empa pooled group and the placebo group with respect to FPG change from baseline (two-sided nominal p-value = 0.01). The empa group on average achieved more weight reduction than the placebo group, though the difference is not statistically significant (two-sided nominal p-value = 0.05).

Table 12: Fasting Plasa Glucose (mg/dL) Change from Baseline at Week 26

	Empa pooled N=52	Placebo N=53
Baseline, mean (SD)	154.43 (57.78)	158.62 (53.80)
Change from baseline, LSMean[1] (SE)	-18.60 (9.28)	17.08 (8.74)
Difference from Placebo, LSMean[1] (CI)	-35.68 (-60.67, -10.70)	
Nominal two-sided p-value	0.01	

Abbreviations: CI = confidence interval, SD = standard deviation, SE = standard error.
[1] The LSMean estimate was based on an ANCOVA model adjusted for baseline FPG, baseline age stratum (< 15 years vs 15 to <18 years), and treatment. Missing data was multiply imputed based on the method of placebo washout. Inference results were combined with Rubin's Rule.
Source Sponsor's analysis; adfpg.xpt, and adsl.xpt

Table 13: BMI Z-Score Change from Baseline at Week 26

	Empa pooled N=52	Placebo N=53
Baseline, mean (SD)	2.95 (0.83)	2.90 (0.99)
Change from baseline, LSMean[1] (SE)	-0.09 (0.03)	-0.01 (0.03)
Difference from Placebo, LSMean[1] (CI)	-0.08 (-0.16, 0.00)	
Nominal two-sided p-value	0.05	

Abbreviations: CI = confidence interval, SD = standard deviation, SE = standard error.
[1] The LSMean estimate was based on an ANCOVA model adjusted for baseline BMI Z-score, baseline age stratum (< 15 years vs 15 to <18 years), and treatment. Missing data was multiply imputed based on the method of placebo washout. Inference results were combined with Rubin's Rule.
Source Sponsor's analysis; adsl.xpt, advs.xpt, and adsl.xpt

[6] Results in Tables 12 and 13 were based on Sponsor's IR response dated March 28, 2023 as per Agency's request for updated analyses for FPG using MI based on the placebo-washout method and an analysis of BMI Z-score. In the original submission, the FPG data were analyzed with missing data imputed based on Last Observation Carried Forward. Weight was analyzed on the original scale (instead of BMI Z-score).

3.3 Evaluation of Safety

Hypoglycemic events were evaluated among the safety set, defined as all subjects who received at least one dose of the treatment. Subjects were analyzed according to their assigned treatments: empa pooled, vs. placebo, from Week 0 to Week 26. No severe hypoglycemia events were observed in the study. The results for hypoglycemia events with Plasma Glucose (PG) < 54 mg/dL (Level 2 hypoglycemia) and for hypoglycemia events with PG ≤ 70 mg/dL (any hypoglycemia) are presented in Tables 14 and 15, respectively[7]. Compared to the placebo, subjects treated with empa showed an elevated risk for both Level 2 hypoglycemia (risk ratio = 2.73, two-sided p-value = 0.16) and for any hypoglycemia event (risk ratio = 1.89, two-sided p-value = 0.30).

Table 14: Analysis of hypoglycemia (PG < 54 mg/dL) up to Week 26, Treated Set

	Empa pooled N = 52	**Placebo N = 53**
Incidence (%)	10 (19.2)	4 (7.5)
Number of events	21	8
Total time at risk (patient year)	23.90	25.08
Unadjusted event rate	0.88	0.32
Adjusted event rate[1], events per patient year (95% CI)	0.86 (0.34, 2.18)	0.31 (0.11, 0.91)
Comparison vs. placebo Adjusted event rate ratio[1] (95% CI)		2.73 (0.67, 11.20)
Nominal p-value (two-sided)		0.16

Abbreviations: CI = confidence interval

[1] The adjusted event rate and rate ratio were based on a negative binomial regression, adjusted for treatment and age stratum (< 15 years vs 15 to <18 years), and offset by time of exposure to treatment.

Source Sponsor's analysis; adsl.xpt, adae.xpt, and adhypo.xpt

Table 15: Analysis of hypoglycemia (PG ≤ 70 mg/dL) up to Week 26, Treated Set

	Empa pooled N = 52	**Placebo N = 53**
Incidence (%)	15 (28.8)	7 (13.2)
Number of events	69	42
Total time at risk (patient year)	23.90	25.08
Unadjusted event rate, events per patient year	2.89	1.67
Adjusted event rate[1], events per patient year (95% CI)	2.86 (1.23, 6.61)	1.51 (0.64, 3.53)
Comparison vs. placebo Adjusted event rate ratio[1] (95% CI)		1.89 (0.57, 6.33)
Nominal p-value (two-sided)		0.30

Abbreviations: CI = confidence interval

[1] The adjusted event rate and rate ratio were based on a negative binomial regression, adjusted for treatment and age stratum (< 15 years vs 15 to <18 years), and offset by time of exposure to treatment.

Source Sponsor's analysis; adsl.xpt, adae.xpt, and adhypo.xpt

[7] Results in Tables 14 and 15 were based on Sponsor's IR response dated April 27, 2023 as per Agency's request of post-hoc analyses of hypoglycemia event counts with negative binomial regression models, adjusted for relevant covariates and offset by exposure time. In the original submission, hypoglycemia events were analyzed descriptively.

According to the clinical reviewer, Dr. Kim Shimy, previous adult studies have found an elevated risk of hypoglycemia when empa is used concomitantly with insulin (see Warnings and Precautions in Section 5.5 of the Jardiance Label). In this study, 25 subjects (48%) treated with empa and 21 subjects (40%) treated with placebo were on background insulin treatment. Tables 16 and 17 display the results of subgroup analyses on hypoglycemia events (PG < 54 mg/dL) based on background insulin use (Yes vs. No). Different from the adult studies, an increased risk of hypoglycemia was observed in both insulin and non-insulin groups in this pediatric study.

Table 16: Analysis of hypoglycemia (PG < 54 mg/dL) up to Week 26, in Subjects on background insulin treatment

	Empa pooled (N = 25)	**Placebo (N = 21)**
Incidence (%)	6 (24.0)	3 (14.3)
Number of events	11	4
Total time at risk (patient year)	11.35	9.74
Unadjusted event rate	0.97	0.41
Adjusted event rate[1], events per patient year (95% CI)	1.07 (0.36, 3.16)	0.27 (0.06, 1.18)
Comparison vs. placebo Adjusted event rate ratio[1] (95% CI)		3.89 (0.59, 25.57)
nominal p-value (two-sided)		0.16

Abbreviations: CI = confidence interval

[1] The adjusted event rate and rate ratio were based on a negative binomial regression, adjusted for treatment and age stratum (< 15 years vs 15 to <18 years), and offset by time of exposure to treatment.

Source reviewer's analysis; adsl.xpt, adae.xpt, and adhypo.xpt

Table 17: Analysis of hypoglycemia (PG < 54 mg/dL) up to Week 26, in Subjects not on background insulin treatment

	Empa pooled (N = 27)	**Placebo (N = 32)**
Incidence (%)	4 (14.8)	1 (3.1)
Number of events	10	4
Total time at risk (patient year)	12.55	15.35
Unadjusted event rate	0.80	0.26
Adjusted event rate[1], events per patient year (95% CI)	0.65 (0.14, 3.04)	0.15 (0.03, 0.89)
Comparison vs. placebo Adjusted event rate ratio[1] (95% CI)		4.25 (0.41, 43.62)
p-value (two-sided)		0.22

Abbreviations: CI = confidence interval

[1] The adjusted event rate and rate ratio were based on a negative binomial regression, adjusted for treatment and age stratum (< 15 years vs 15 to <18 years), and offset by time of exposure to treatment.

Source reviewer's analysis; adsl.xpt, adae.xpt, and adhypo.xpt

4 FINDINGS IN SPECIAL/SUBGROUP POPULATIONS

4.1 Sex, Race, Age, and Geographic Region

Subgroup analyses on HbA1c (%) change from baseline at Week 26 were conducted with respect to the baseline characteristics: sex, race, age (< 15 years, and 15 to < 18 years), and region[8] (US vs outside of US). Each analysis modeled the primary endpoint with an ANCOVA adjusted for baseline HbA1c, treatment, age stratum at randomization (except for the subgroup analysis on age), subgroup and subgroup-by-treatment interaction. Similar to the primary efficacy analysis, missing data were multiply imputed based on placebo washout and the inference results were combined via Rubin's Rule.

Additionally, the Bayesian shrinkage analyses based on the sample estimates were performed. For a given baseline characteristic (e.g., sex), when estimating the treatment effect within a subgroup (e.g., the male subgroup), the shrinkage method borrows information from the other subgroup(s) (the female subgroup), and hence is considered a "weighted" average of the sample estimate and the overall estimate. The weights are based on the ratio of the between-subgroup variability to the within-subgroup variability. A small ratio indicates a small between-subgroup variability relative to the within-subgroup variability. Consequently, more weight is put on the overall estimate, and more shrinkage is applied.

For a given baseline characteristic with *k* subgroups, let Y_i (i = 1, ... k) be the observed sample estimate of the treatment effect in subgroup *i*. The shrinkage analysis in this review assumes the following:

- $Y_i \sim N(\mu_i, \sigma_i^2)$, where μ_i is the expected treatment effect for subgroup *i*, and σ_i^2 is the within-subgroup variance
- σ_i^2 is set to the variance for the sample estimate
- $\mu_i \sim N(\mu, \tau^2)$, where $\mu \sim N(0, (6.8)^2)$, and $1/\tau^2 \sim$ Gamma (0.001, 0.001)

The last assumption stated that the expected treatment effect for all *k* subgroups share a common normal distribution centered at μ and with variance τ^2. A non-informative prior, as specified above, was applied to this normal distribution. A standard deviation of 6.8 was chosen for the centrality parameter μ, so that its standard deviation was approximately four times the subject-level standard deviation, which was estimated to be around 1.7 based on the primary analysis results.

The sample estimates and the shrinkage estimates of the treatment difference with respect to HbA1c change from baseline at Week 26 are presented in Figure 3. The point estimates for all subgroup levels were covered by the 95% confidence interval of the treatment effect estimate of the overall population, suggesting homogeneous treatment effects of empa across different subpopulations. Compared to the frequentist's sample estimate, the shrinkage estimate had less variability and a magnitude closer to the overall estimate. This shrinkage effect on sample estimates was particularly strong for subgroup levels that have small sample sizes and large variability (e.g., the race category *Other*).

[8] The variable "region" was derived based on countries' names.

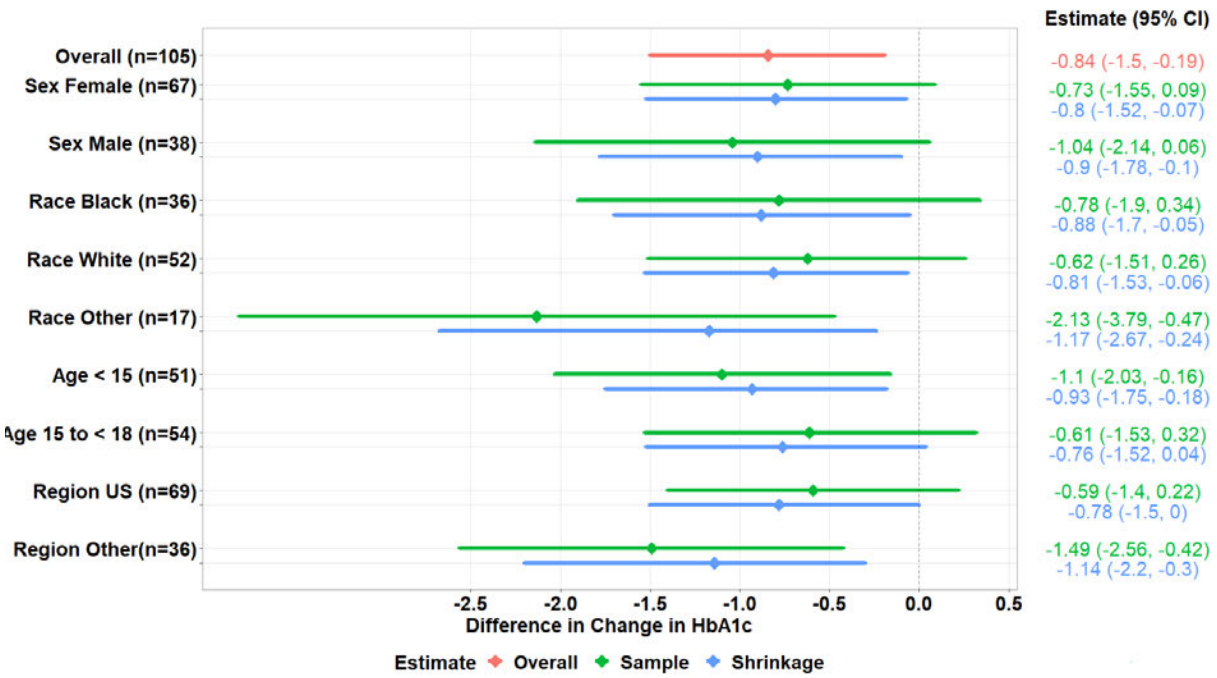


Figure 3: Placebo-Adjusted HbA1c Change from Baseline, Subgroup Analyses
Source reviewer's analysis; ada1c,xpt, adsl.xpt

When performing the subgroup analysis on race, the race categories American Indian/Alaska Native (n = 5), Asian (n = 5), Multiple (n = 5), Native Hawaiian/ Other Pacific Islander (n = 1), as well as one subject with missing race information, were combined into the race category "Other", due to insufficient sample sizes. Compared to the race categories Black and White, an uncommonly large treatment effect difference was observed in this "Other" category. For descriptive purpose, Figure 4 displays the treatment effect for each subject from the "Other" category. Specifically, the dots colored by race and aligned by treatment arms represent the observed primary endpoint values (i.e., HbA1c change from baseline at Week 26) for individuals from the "Other" category. The two vertical grey bars are the 95% confidence intervals of the treatment effect for the two treatment arms estimated based on the primary efficacy model. Only two observations from the "multiple" racial category and treated with empa are covered by the confidence interval, whereas the rest of the observations are beyond the coverage. The uncommonly large treatment effect in the "Other" category seems to be driven by outliers both on the low end from the empa arm, and on the high end from the placebo arm.

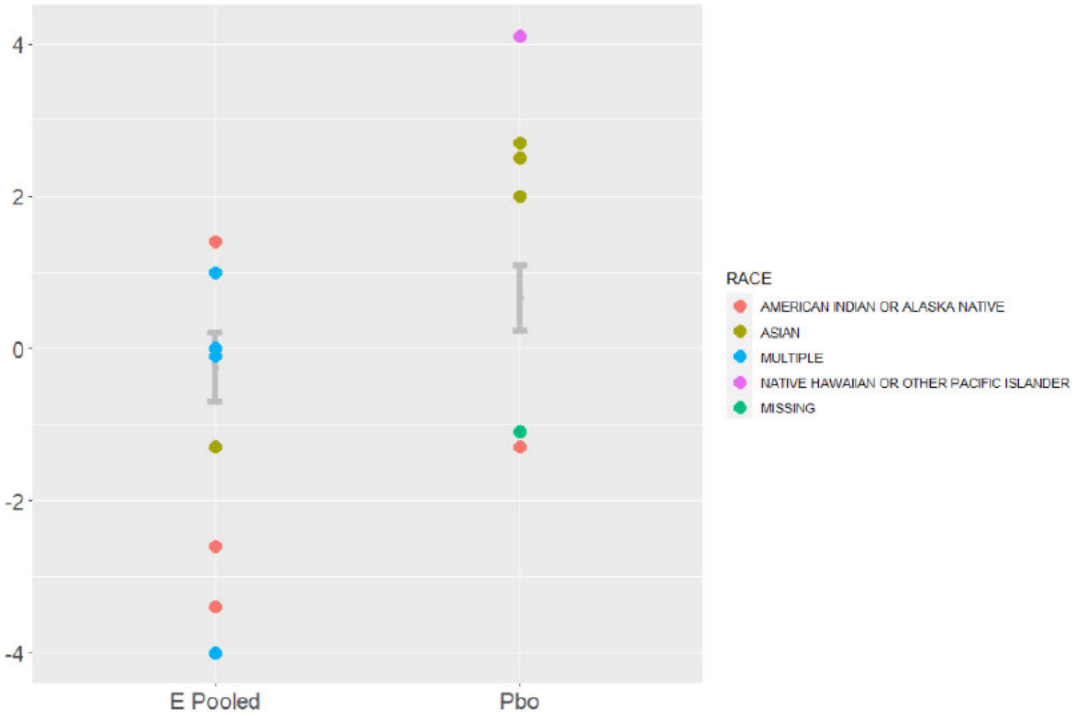


Figure 4: HbA1c Change at Week 26 from Baseline, A Breakdown of the Race Category *Other*
Source reviewer's analysis; ada1c,xpt, adsl.xpt

4.2 Other Special/Subgroup Populations

Subgroup Analysis Based on Background Therapies

Subgroup analyses on background medications were performed to examine the treatment effect of empa in combination with metformin. Subjects in this study were on one of the four background therapy regimens (n, %): metformin only [met only] (54, 51.4%), metformin and insulin [met + insulin] (41, 39.0%), insulin only (5, 4.8%), or none (5, 4.8%). Figure 5 presents the results of three analyses, each based on a different grouping of background medications as follows:

- bkgrd1
  - met only
  - met + insulin
  - the other (including "insulin only" and "none")
- bkgrd2
  - met (including "met only" and "met + insulin")
  - the other (including "insulin only" and "none")
- bkgrd3
  - met only
  - the other (including "met + insulin", "insulin only" and "none")

The grouping *bkgrd1* intends to compare across all different background regimens. "Insulin only" and "none" were combined into the category "the other" due to insufficient sample sizes.

Bkgrd 2 focuses on the comparison of metformin use vs. no metformin use. *Bkgrd3* focuses on the comparison of metformin as monotherapy vs. the other.

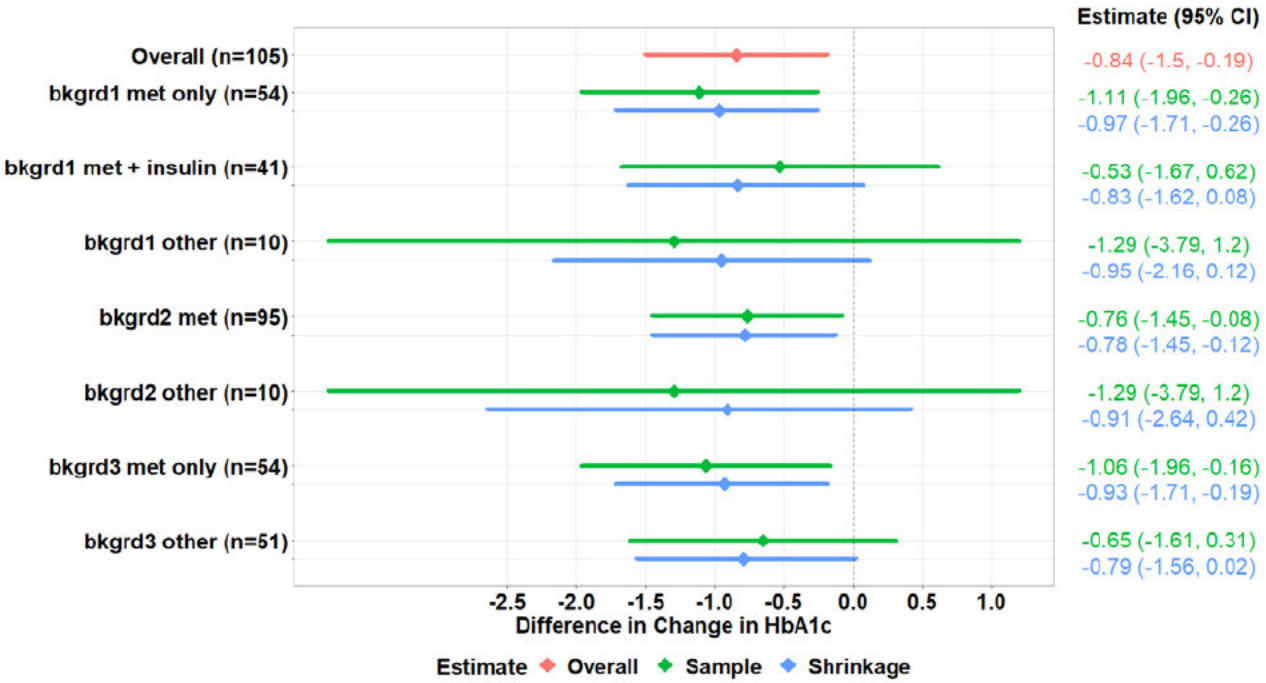


Figure 5: HbA1c Change from Baseline at Week 26, Subgroup Analysis on Background Therapy
Source reviewer's analysis; ada1c,xpt, adsl.xpt

As shown in Figure 5, the estimated treatment effects for all subgroup levels based on different groupings were consistent with the overall population. Further, as the majority of the study participants were treated with metformin (91.4 %), the confidence interval based on the full analysis set highly overlaps with the confidence interval based on subjects on metformin therapy (i.e., bkgrd2 met).

Baseline HbA1c as an Effect Modifier

It is well known that baseline HbA1c is an effect modifier; i.e., the treatment effect measured by A1c change from baseline depends on a patient's baseline HbA1c measurement. Figure 6 below is a scatter plot of HbA1c change from baseline at Week 26 vs. baseline HbA1c. The scatter points are color-coded by treatment arms. Two regression lines based on completers from empa and placebo are superimposed over the scatter points. The regression line is $y = -0.49\,x + 3.63$ for empa, and $y = 0.35x - 2.30$ for placebo. The difference in slopes is 0.84, which implies that for every 1% increase in baseline HbA1c, the placebo-adjusted treatment effect measured by HbA1c change from baseline increases by 0.84%. As an illustration, when baseline HbA1c is 8%, the average change from baseline for empa and placebo are -0.29% and 0.5%, respectively, which amounts to a difference of 0.79%. However, when baseline HbA1c is 9%, the average change from baseline for empa and placebo are -0.78% and 0.85%, which amounts to a difference of 1.63%. The higher the baseline HbA1c, the larger the treatment effect. In the primary analysis, baseline HbA1c was included in the ANCOVA model to adjust for this modification effect.

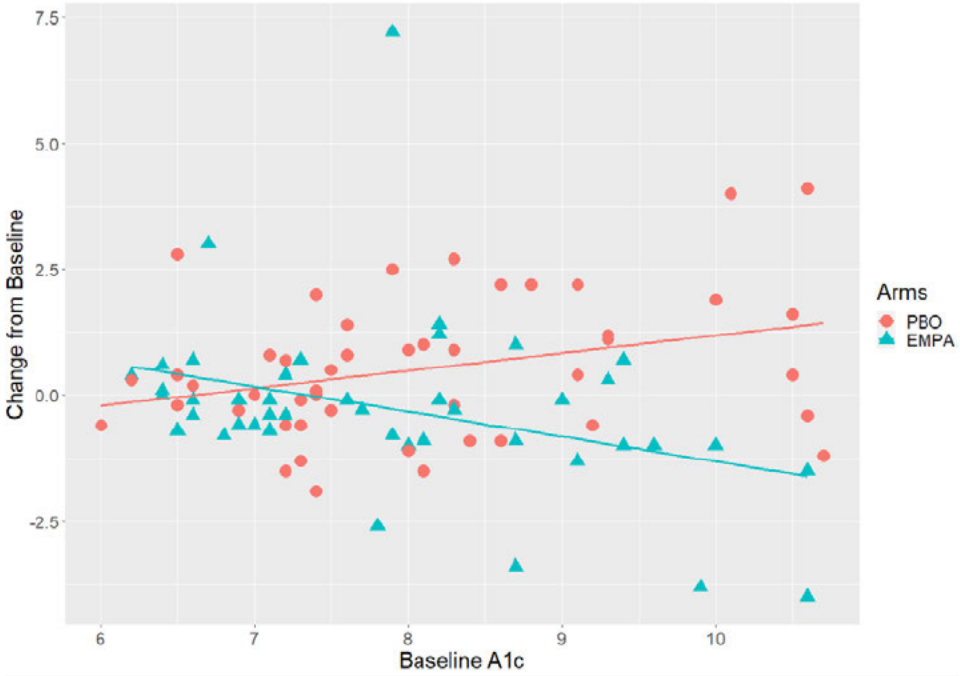


Figure 6: Scatterplot of Baseline A1c vs. Change from Baseline
Source reviewer's analysis; ada1c,xpt, adsl.xpt

5 SUMMARY AND CONCLUSIONS

5.1 Statistical Issues

No major statistical issues were identified in this review. The overall missing rate for the primary endpoint measurements was 9.6% for empa, and 5.7% for placebo. Missing endpoints were multiply imputed based on placebo washout. For primary efficacy analyses, the applicant applied an ANCOVA adjusted for treatment (placebo, empa or lina), baseline HbA1c and age stratum at baseline (< 15 years vs 15 to <18 years).

In this review, minor review issues include an inefficient testing structure for the primary and secondary hypothesis families, and a lack of dose-response for the empa 25 mg and 10mg (addressed in Section 3.2.4). No changes on the efficacy conclusion were identified.

5.2 Collective Evidence

For the primary efficacy analysis, the placebo-adjusted treatment effect for empa with respect to HbA1c change from baseline at Week 26 was -0.84 with a 95% CI (-1.50, -0.19) and a two-sided p-value = 0.01. Additionally, results from sensitivity analyses confirmed robustness of the primary efficacy results to untestable assumptions on missing data. Subgroup analyses on the primary efficacy endpoint found consistent treatment effect of empa in subgroup levels based on

age, sex, race, and region. as well as background medications. A numerically elevated risk of hypoglycemia was found in subjects treated with empa compared to those treated with placebo.

5.3 Conclusion and Recommendations

Statistical analyses based on the clinical data from the Phase 3 pediatric study DINAMO have demonstrated robust evidence to support the effectiveness of empagliflozin regarding glycemic control among pediatric patients (11 to < 18 years) with T2DM. The statistical review team recommend approval of the proposed label updates for Jardiance, Synjardy and Synjardy XR.

5.4 Labeling Recommendations

Figure 7 displays the proposed change for *Section 8.4: Pediatric Use* of the current label.

8.4 Pediatric Use

The safety and effectiveness of JARDIANCE as an adjunct to diet and exercise to improve glycemic control in type 2 diabetes mellitus have (b) (4) been established in pediatric patients aged 10 years and older. Use of JARDIANCE for this indication is supported by evidence from a 26-week double-blind, placebo-controlled clinical trial, with a double-blind active treatment safety extension period of up to 52 weeks in 157 pediatric patients aged 10 to 17 years with type 2 diabetes and a pediatric pharmacokinetic study *[see Clinical Pharmacology (12.3) and Clinical Studies (14)]*.

The safety and effectiveness of JARDIANCE have not been established in pediatric patients less than 10 years of age.

Figure 7: Proposed Label
Source: proposed uspi by BIPI

In support of this pediatric indication, a new section on pediatric clinical studies (*Section 14.2 Glycemic Control Trial in Pediatric Patients Aged 10 to 17 years with Type 2 Diabetes Mellitus*) was added to Section 14 of the product label. Table (b) (4) in Section 14.2 presents the analysis results on the primary endpoint HbA1c change from baseline at Week 26, and the secondary endpoint FPG change from baseline at Week 26. During the recent labelling meeting, we requested several edits for Table (b) (4) as follows:

- Replace (b) (4) with results based on MI imputation with placebo-washout.
- Add the definition of mITT population in the table footnote
- Describe details about the MI method for the primary efficacy analysis
- Add a footnote on the missing data rate for the study
- Clarify if the p-value presented in the table is two-sided or one-sided

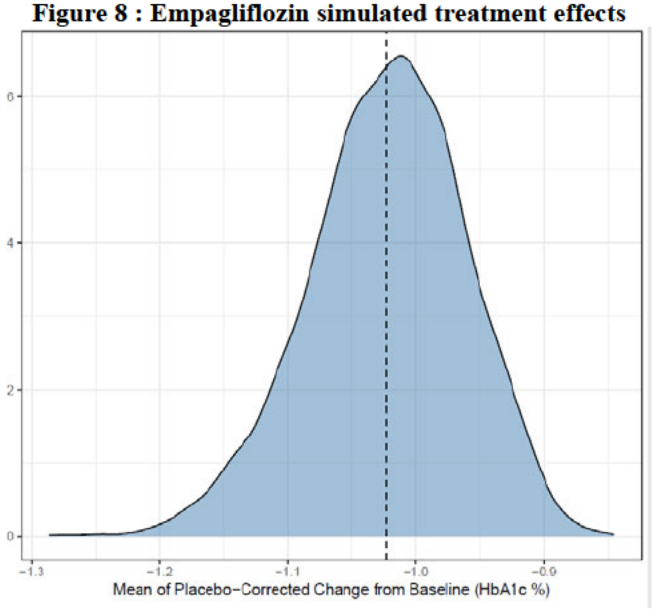
6 Appendix: Supplementary Bayesian Analysis

6.1 Statistical Methodologies

Bayesian inference was used in this supplementary analysis to leverage information from the previously fitted pharmacokinetic and exposure-response models for empagliflozin and based on available historical data in adult and pediatric patients with T2DM. Partial exchangeability of pediatric and adult data through covariate adjustment. The objective of the Bayesian analysis was to provide supportive evidence for the comparison of the mean change in HbA1c (%) from baseline to the end of 26 weeks between empagliflozin and placebo. Throughout the review we will refer to this as the placebo-corrected treatment effect of linagliptin. Bayesian borrowing analyses based on two prior approaches were conducted.

Prior based on exposure-response based pharmacometric model: Using the available data from the completed studies, the applicant constructed an exposure-response based pharmacometric model. Using this model, the applicant generated 5000 simulation-predicted treatment responses for HbA1c based on the parameters observed in the pediatric population to serve as the prior distribution for the Bayesian borrowing analysis. The sample means from each of the 5,000 corresponding iterations in the pharmacometric simulation constitute a random sample of the predicted placebo-corrected treatment effect in DINAMO. Due to the nature of the pharmacometric models, it is assumed that the predicted placebo-corrected mean HbA1c change from baseline approximately follows a normal distribution (Figure 8). This approximate normal distribution had a mean of μ_I=−1.02% and a standard deviation of 0.06%.

Figure 8 : Empagliflozin simulated treatment effects



Source: Excerpted from page 36 of the supplemental analysis report

The unit standard deviation (SD), σ_I,was estimated from the clinical trials in adults and the blinded assessment of the DINAMO study. The estimated SD of empagliflozin, linagliptin, and placebo arm were 1.52, 1.65 and 1.72 respectively. Assuming mutual independence among the

treatment arms, the SD corresponding to the empagliflozin treatment difference was found to be $\sigma_I = \sqrt{1.52^2 + 1.72^2} = 2.29$. In order to obtain comparisons that correspond to an informative prior weight of at most 100 patients per treatment group, the applicant replaced the prior variances v_I from the pharmacometrics simulations with:

$$v_I^* = \begin{cases} v_I, & \text{if } v_I \geq \sigma_I^2/100 \\ \dfrac{\sigma_I^2}{100}, & else \end{cases}$$

The informative prior is then robustified against potential prior-data conflict. The final prior is a mixture of the informative prior part with weight w_I= 0.65 and a weakly informative normally distributed prior component with mean μ_I corresponding to the mean placebo-corrected treatment effect estimated from the pharmacometric simulation, and standard deviation σ_I. This leads to the following final prior probability densities for the placebo corrected treatment effect:

$$p_I(\theta_I) = w_I Norm(\mu_I, v_I^*) + (1 - w_I) Norm(\mu_I, \sigma_I^2).$$

The weight w_I=0.65 was determined to have overall prior ESS of 100 in the linagliptin and placebo arm combined.

Prior based on data from other SGLT-2 inhibitors: A second Bayesian borrowing analysis to leverage prior data in a blinded assessment of DINAMO was performed. In this analysis the placebo-corrected treatment effects reported for pediatric populations with T2DM treated with dapagliflozin was considered to be informative for the outcome in DINAMO. One study with "Farxiga 10 mg" (dapagliflozin) was identified for providing prior information about the placebo-corrected treatment effects in DINAMO (Table 18).

Table 18: HbA1c (%) change in historic placebo-controlled trials of SGLT-2 or DPP-4 inhibitors in pediatric patients with T2DM

	Placebo			Active treatment			Treatment difference[4]
Active treatment / Endpoint	N	Mean	Variabi-lity	N	Mean	Variabi-lity	Adj. mean active – placebo (95% CI)
Farxiga 10mg[1]							
Treated patients	33			39			
HbA1c change at Week 24	23	0.50	SE=0.34	31	-0.25	SE=0.30	-0.75 (-1.65, 0.15)
Januvia 100mg[2]							
HbA1c change at Week 20	95	0.23		95	0.06		-0.17 (-0.62, 0.28)
Januvia 100mg[3]							
HbA1c change at Week 20	113	0.09		107	-0.23		-0.33 (-0.70, 0.05)

Sources: Page 34 of the Statistical Analysis Plan for the Bayesian borrowing Analysis.
[1] Study to Evaluate Safety and Efficacy of Dapagliflozin in Patients with Type 2 Diabetes Mellitus Aged 10-24 Years ClinicalTrials.gov Identifier: NCT02725593
[2] US PI for Januvia, revised 12/2020, Section 8.4 Pediatric use, study 1
[3] US PI for Januvia, revised 12/2020, Section 8.4 Pediatric use, studies 2 and 3
[4] Treatment difference: Least Square Mean difference of active treatment – placebo.

Based on the reported standard errors for the placebo-corrected effect and the sample sizes for the dapagliflozin and placebo arm the estimated unit-information SD was σ_I=2.33. The informative component of the prior was derived from the adjusted mean difference for Farxiga (-0.75) and unit-information SD 2.33 using the meta analytic predictive (MAP) [Schmidli et al, 2014] approach. Further a moderate prior was used for the between-study heterogeneity (τ). Modelling $\tau \sim Half - Normal\left(scale = \frac{\sigma_I}{16}\right)$[9] , the resulting informative prior for empagliflozin was found to be normally distributed with a mean −0.72 and a standard deviation of 0.48. The weight of the informative prior component is set to 75% following discussion with clinical experts, with the remaining 25% weight allocated to a unit information prior with mean -0.72 to provide robustness in the case of prior-data conflict. This resulted in an effective sample size (ESS) of 14 per treatment arm. The ESS was calculated with the expected local information ratio (ELIR). The final parametric robust MAP prior distribution has the probability density as below approach,

$$p_E(\theta_E) = 0.75\ Norm(-0.72, 0.48^2) + 0.25\ Norm(-0.72, 2.33^2).$$

Posterior calculation: Posterior distributions were derived from each prior and the observed placebo-corrected treatment effect in the DINAMO trial. In this analysis, the comparison of the

[9] If $X \sim N(0, \sigma^2)$ then $|X|$ is said to have Half-Normal distribution with scale=σ.

97.5% quantile of the posterior treatment effect with 0 was used to determine if the analysis indicated superior efficacy of the treatment:

$$\text{Prob}\,(\theta_E < 0 \mid y) \geq 0.975$$

where:

- y is the observed data.
- θ_E is the placebo-corrected effect of empagliflozin.

If these decision criteria were met, then there was evidence of superior efficacy of the treatment in the pediatric population of DINAMO.

The applicant performed a tipping point sensitivity analysis with alternative prior weights of 0, 0.1, 0.2, …, 0.9, 1. Here a weight of 0 corresponded to a weakly-informative prior, with the resulting estimate being based almost entirely on the DINAMO data. A weight of 1 corresponded to a prior that is entirely based on the pharmacometric model predictions (or the literature data on SGLT-2 inhibitors), assuming full exchangeability of the covariate-adjusted predictions (or the literature data on SGLT-2 inhibitors) with the DINAMO outcome data without robust component down-weighting.

6.2 Results and Conclusions

The estimated placebo-corrected treatment effect from the DINAMO study using the pediatric data alone was −0.84% with a standard error of 0.33%, from which the likelihood was derived.

Bayesian borrowing based on exposure-response based pharmacometric model

The prior SD from the pharmacometric simulations (0.06) was less than the threshold for an ESS of 100 ($0.229 = 2.29/\sqrt{100}$). Therefore, the standard deviation of the informative component was set to 0.229. This resulted in the following prior distribution:

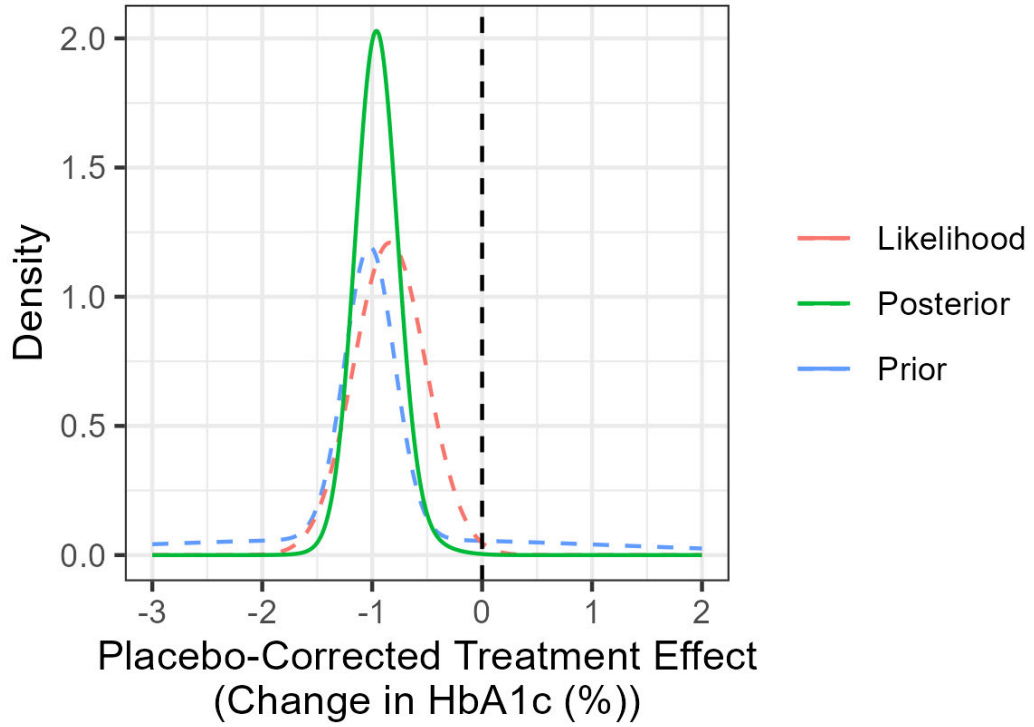
$$p_I(\theta_I) = w_I Norm\left(\mu_I, v_I^*\right) + (1 - w_I) Norm\left(\mu_I, \sigma_I^2\right),$$

where $\mu_I = -1.02$, $v_I^* = 0.229^2$, $\sigma_I^2 = 2.29^2$.

This robust prior distribution had a mean of −1.02% and a standard deviation of 1.37%. This robust prior results in heavier tails, with 2.5% quantile (−4.38%) and 97.5% quantile (2.33%) that are further from the mean than the approximately 2 standard deviations.

Comparison of the prior, likelihood, and posterior distributions of the mean placebo-corrected treatment effect did not show any evidence of prior-data conflict (Figure 9).

Figure 9: Empagliflozin placebo-corrected treatment effect distributions



Source: Statistical Reviewer's Analyses

The posterior mean placebo-corrected treatment effect was −0.95%, with a standard deviation of 0.21%. The 97.5% quantile was −0.53%, which was less than zero corresponding to superior efficacy for empagliflozin compared to placebo. The posterior probability of the placebo-corrected treatment effect being less than zero was greater than 0.99.

The tipping point sensitivity analyses showed that for any choice of prior mixture weight, the 97.5% decision threshold was satisfied (Table 19).

Table 19: Tipping point sensitivity analysis for different prior weights. 0 corresponded to only using the weakly-informative prior and 1 corresponded to only using the pharmacometric simulation results as the prior.

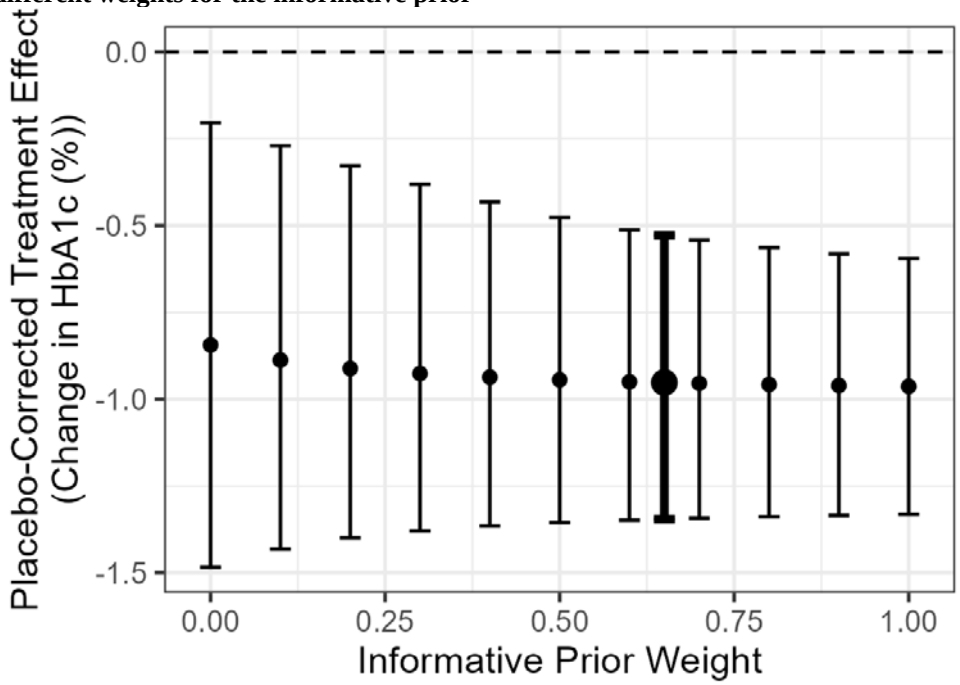
Informative Prior Weight	Prior ESS per Treatment Arm	Posterior Probability of Superior Efficacy	97.5% Decision Rule Met	Posterior Mean Treatment Effect	95% Equal-tailed Credible Interval
0.65	51	>0.999	YES	-0.95	(-1.35, -0.53)
0	1*	0.995	YES	-0.84	(-1.48, -0.20)
0.1	4	0.997	YES	-0.89	(-1.43, -0.27)
0.2	9	0.998	YES	-0.91	(-1.40, -0.33)

0.3	16	0.998	YES	-0.93	(-1.38, -0.38)
0.4	25	0.999	YES	-0.94	(-1.37, -0.43)
0.5	35	0.999	YES	-0.94	(-1.36, -0.48)
0.6	45	0.999	YES	-0.95	(-1.35, -0.51)
0.7	56	>0.999	YES	-0.95	(-1.34, -0.54)
0.8	69	>0.999	YES	-0.96	(-1.34, -0.56)
0.9	83	>0.999	YES	-0.96	(-1.33, -0.58)
1	98	>0.999	YES	-0.96	(-1.33, -0.59)

Source: Statistical Reviewer's Analyses
[*] : With 0 weight to the informative component, the robust component of the prior contributes 1 patient worth of information.

Furthermore, as the informative prior weight increased, the width of the credible intervals decreased, and the mean estimate was closer to the prior mean. This reflected the increased information and lower variability in the informative prior compared to the robust prior component.

Figure 10: Empagliflozin placebo-corrected treatment effects and 95% equal-tailed credible intervals for different weights for the informative prior

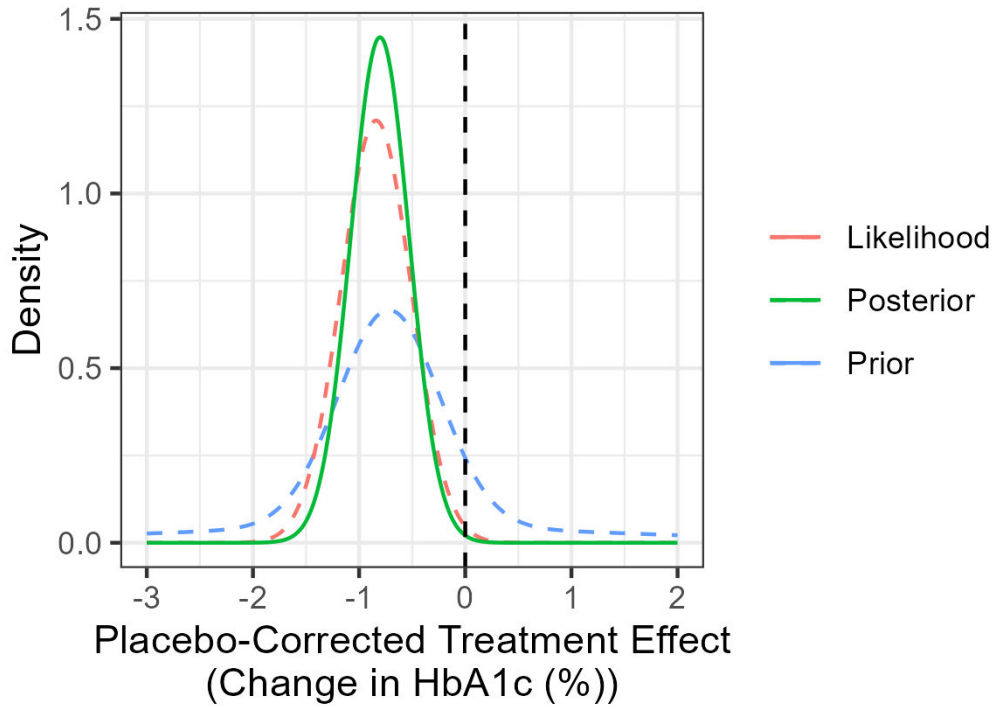


Source: Statistical Reviewer's Analyses
Sensitivity analysis for different prior weights for the informative prior component. A weight of 0 corresponded to only using the weakly-informative prior and 1 corresponded to only using the pharmacometric simulation results as the prior, and the bolded interval is the weight used in the primary analysis. The horizontal dashed line corresponds to the null value (0). Intervals are 95% credible intervals, and points are the posterior mean.

Bayesian borrowing based on data from other SGLT-2 inhibitors

The robust prior distribution, derived from the meta-analysis of SGLT-2 inhibitors, had a mean of −0.72% and a standard deviation of 1.24%. This prior had a smaller effect size and larger standard deviation than the one based on pharmacometric simulations for empagliflozin. The prior, likelihood and the posterior density plots are given in Figure 11.

Figure 11: Empagliflozin placebo-corrected treatment effect distributions



Source: Statistical Reviewer's Analyses

The posterior mean placebo-corrected treatment effect was −0.80%, with a standard deviation of 0.28%. The 97.5% quantile was −0.26%, which was less than zero corresponding to superior efficacy for empagliflozin compared to placebo. The posterior probability of the placebo-corrected treatment effect being less than zero greater than 0.99.

The tipping point sensitivity analyses showed that for any choice of prior mixture weight, the 97.5% decision threshold was satisfied (Table 20). Furthermore, as the informative prior weight increased, the width of the credible intervals (Figure 12) decreased slightly. This reflected the increased information and lower variability in the informative prior compared to the robust prior component.

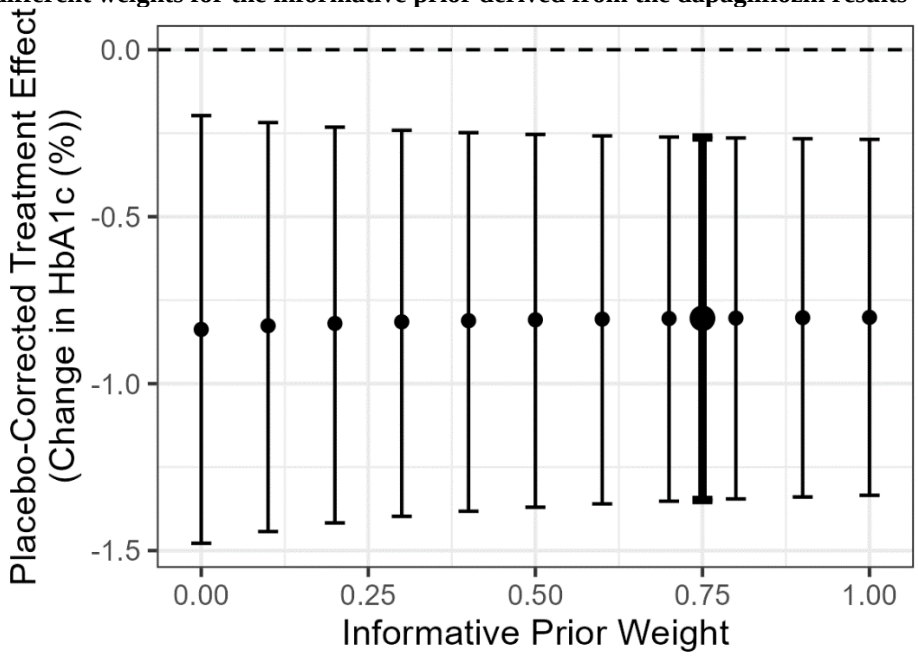
Table 20: Empagliflozin sensitivity analysis (based on dapagliflozin)

Informative Prior Weight	Prior ESS per Treatment Arm	Posterior Probability of	97.5% Decision Rule Met	Posterior Mean	95% Equal-tailed Credible Interval

		Superior Efficacy		Treatment Effect	
0.75	14	0.998	YES	-0.80	(-1.35, -0.26)
0	1*	0.995	YES	-0.84	(-1.48, -0.20)
0.1	1	0.996	YES	-0.83	(-1.44, -0.22)
0.2	2	0.997	YES	-0.82	(-1.42, -0.23)
0.3	4	0.997	YES	-0.81	(-1.40, -0.24)
0.4	5	0.997	YES	-0.81	(-1.38, -0.25)
0.5	7	0.998	YES	-0.81	(-1.37, -0.25)
0.6	10	0.998	YES	-0.81	(-1.36, -0.26)
0.7	12	0.998	YES	-0.81	(-1.35, -0.26)
0.8	15	0.998	YES	-0.80	(-1.35, -0.26)
0.9	19	0.998	YES	-0.80	(-1.34, -0.27)
1	24	0.998	YES	-0.80	(-1.33, -0.27)

Source: Statistical Reviewer's Analyses.
* : With 0 weight to the informative component, the robust component of the prior contributes 1 patient worth of information.

Figure 12: Empagliflozin placebo-corrected treatment effects and 95% equal-tailed credible intervals for different weights for the informative prior derived from the dapagliflozin results



Source: Statistical Reviewer's Analyses
Sensitivity analysis for different prior weights for the informative prior component. A weight of 0 corresponded to only using the weakly-informative prior and 1 corresponded to only using the pediatric trial result from dapagliflozin as the prior. The horizontal dashed line corresponds to the null value (0). The bolded interval is the weight used in the primary analysis. Intervals are 95% credible intervals.

In conclusion, the Bayesian borrowing analyses based on both the pharmacometric simulation and literature data from other SGLT-2 inhibitors support superior efficacy for empagliflozin compared to placebo. Evidence of superior efficacy was obtained across all alternative weight to the informative component which was consistent with the primary efficacy analysis.

This is a representation of an electronic record that was signed electronically. Following this are manifestations of any and all electronic signatures for this electronic record.

/s/

WENDA TU
05/26/2023 12:41:14 PM

SATYAJIT GHOSH
05/26/2023 12:43:14 PM

JAMES E TRAVIS
05/26/2023 12:46:19 PM
I concur

YOONHEE KIM
05/26/2023 12:48:08 PM
I concur.