Application of Adaptive Design Methodology in Development of a Long-Acting Glucagon-Like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Zachary Skrivanek, Ph.D.,[1] Scott Berry, Ph.D,[2] Don Berry, Ph.D.,[2,3] Jenny Chien, Ph.D.,[1] Mary Jane Geiger, M.D., Ph.D.,[1] James H. Anderson, Jr., M.D.,[4] and Brenda Gaydos, Ph.D.[5]

Abstract

Background:

Dulaglutide (dula, LY2189265), a long-acting glucagon-like peptide-1 analog, is being developed to treat type 2 diabetes mellitus.

Methods:

To foster the development of dula, we designed a two-stage adaptive, dose-finding, inferentially seamless phase 2/3 study. The Bayesian theoretical framework is used to adaptively randomize patients in stage 1 to 7 dula doses and, at the decision point, to either stop for futility or to select up to 2 dula doses for stage 2. After dose selection, patients continue to be randomized to the selected dula doses or comparator arms. Data from patients assigned the selected doses will be pooled across both stages and analyzed with an analysis of covariance model, using baseline hemoglobin A1c and country as covariates. The operating characteristics of the trial were assessed by extensive simulation studies.

Results:

Simulations demonstrated that the adaptive design would identify the correct doses 88% of the time, compared to as low as 6% for a fixed-dose design (the latter value based on frequentist decision rules analogous to the Bayesian decision rules for adaptive design).

Conclusions:

This article discusses the decision rules used to select the dula dose(s); the mathematical details of the adaptive algorithm—including a description of the clinical utility index used to mathematically quantify the desirability of a dose based on safety and efficacy measurements; and a description of the simulation process and results that quantify the operating characteristics of the design.

J Diabetes Sci Technol 2012;6(6):1305-1318

Author Affiliations: [1]Lilly Diabetes, Eli Lilly and Company, Indianapolis, Indiana; [2]Berry Consultants, Austin, Texas; [3]The University of Texas M.D. Anderson Cancer Center, Houston, Texas; [4]Symcopeia Company, Carmel, Indiana; and [5]Lilly Research Laboratories, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, Indiana

Corresponding Author: Zachary Skrivanek, Ph.D., Lilly Diabetes, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46285; email address *skrivanek_zachary@lilly.com*

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

Introduction

Dose selection is a pivotal milestone in drug development and has important implications for the ultimate usefulness of a drug. In a typical development program, data available at the end of phase 2 are reviewed and discussed with regulators, and a decision is made about which dose or doses to study in phase 3. In the development program of dulaglutide (dula, LY2189265), a once-weekly glucagon-like peptide-1 (GLP-1) analog in development for the treatment of type 2 diabetes mellitus (T2DM), dose selection, and dose confirmation will be combined into a single adaptive, inferentially seamless trial. (Additional details of this study, entitled "A Study of LY2189265 Compared to Sitagliptin in Patients With Type 2 Diabetes Mellitus on Metformin," can be found at http://clinicaltrials.gov as NCT00734474.)

Dose selection will be based on prespecified rules in the adaptive algorithm. The study team prespecified criteria used to select doses to measure the value of the different doses—the characteristics needed for a dose to be considered efficacious, safe, and competitive in the expected marketplace. These criteria also used to adapt treatment allocation.

Geiger and coauthors[1] describe the rationale and final study design for an adaptive, dose-finding, inferentially seamless phase 2/3 study known as the Assessment of Weekly AdministRation of LY2189265 in Diabetes 5 (AWARD-5), applied in the development of dula. There are three main features that make this design adaptive. First, the probability of a new patient being assigned to a given dose of dula will change, or adapt, based on accumulating data in stage 1. AWARD-5 is divided into two stages that are based on two randomization schemes: an adaptive scheme (stage 1) and a fixed scheme (stage 2).[1] Four response measures based on expert opinion, regulatory feedback, and scientific understanding of the GLP-1 analog class of therapeutics and experience with dula were identified as potentially dose limiting. These measures encompassed both safety and efficacy and together will be combined into a single measure known as the clinical utility index (CUI), which will be used in the adaptive treatment algorithm. Second, dula dose selection for stage 2 will be determined in stage 1 based on accumulating data. Third, sample sizes in stage 1 and stage 2 will be determined adaptively and inform the decision as to how much data will be sufficient to draw conclusions. The design is *inferentially seamless* (i.e., final analysis and conclusions will combine data

from patients enrolled in both stage 1 and stage 2). The final analysis to support advancement to phase 3 will use frequentist statistical methods. The adaptive algorithm employs Bayesian methods. This article discusses the mathematical details of the adaptive algorithm, including a description of the CUI, the decision rules used to select the dula dose(s), and a description of the simulation process and results that quantify the operating characteristics of the design.

Methods

This study was developed as a two-stage design. In the first stage of the trial, a Bayesian framework is used to adaptively allocate patients to 7 doses of dula and to assess decision rules. Patients are allocated with equal probability to the placebo and active comparator arms. We will evaluate two decision rules after 200 subjects were randomized: (1) to stop for futility, based on both safety and efficacy; or (2) to start stage 2 with up to two doses selected from stage 1, based on predefined decision rules. If there is insufficient evidence to make either of these decisions, patients continue to be randomized in stage 1. If sufficient evidence cannot be gathered to make either decision after 400 patients are enrolled, the study will be terminated. If dose selection does occur, stage 2 will commence and will use fixed randomization and a fixed sample size, based on a predictive probability calculation of meeting the study objectives based on stage 1 data. The additional patients from stage 2 will enable further characterization of the safety and efficacy of the selected dula doses. A fundamental difference between the two stages is the randomization scheme. In stage 2, randomization proportions are fixed. To maintain control of a type I error, design implementation could not deviate from the prespecified algorithm unless the Data Monitoring Committee (DMC) stopped the doses for safety considerations.

Dose-Response Measures

As described in Geiger and coauthors,[1] the following four response measures of efficacy and safety were chosen: hemoglobin A1c (HbA1c), weight, heart rate (HR), and diastolic blood pressure (DBP).

For each of the four measures, normal dynamic linear models (NDLM) were used to model the dose response for each dula dose.[2,3] A nonparametric approach to model correlated data, NDLM "borrows" information

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

from neighboring doses but does not force any particular shape to the overall response curve. The dose-response model for HbA1c is modeled at 12 months and at 6 months for other measures. The same measures for placebo and sitagliptin were modeled distinctly from NDLM, with no "borrowing" from other therapies. A normal prior distribution is adopted for the mean end point for each measurement and dose. (See section on Bayesian Methods and Decision Rules for an explanation of prior distributions.)

Clinical Utility Index

Every therapy has benefits and risks. The relative importance of these characteristics depends on the disease, the patient, and how the drug is used. The CUI is a means of weighting and quantifying these tradeoffs and provides a single metric for multiple dimensions of benefit and risk.[4]

The CUI for this study was based on the aforementioned four response measures. The goal of adaptation is to optimize patient exposure to those dula doses demonstrating the best balance of efficacy (HbA1c), desirability (weight reduction), and minimization of potential cardiovascular safety concerns (HR and DBP). To do this, a multiplicative CUI was developed, i.e., a mathematical formula in which these four response measures are differentially quantified and transformed into a single utility index value for each dose. Whenever an adaptation is made in stage 1, all available data from these four response measures will be used to make inferences for the following outcomes: mean 12-month change from baseline in HbA1c relative to sitagliptin and mean 6-month change from baseline in weight, HR, and DBP relative to placebo. In short, the CUI is a weighting of these four end points, which are well accepted as important aspects in diabetes management. Patients were adaptively randomized according to the CUI, and the decision rules were based on the CUI. Uncertainty is associated with evaluating the utility of a dose because evaluation is based on observed data that has inherent variability; therefore, decision rules are implemented with probability statements around the utility.

In constructing the CUI, individual component utility functions for each of the measures were derived based on input from the entire design team and from external consultants, including regulatory authorities. The component utility for HbA1c is anchored so that a minimal clinically acceptable decrease has a value of 1. For each of the other measures, the component function is anchored at a value of 1 for neutrality (i.e., any end point that has a neutral

effect on benefit assessment maps to 1). A value greater than 1 signifies added benefit, and a value less than 1 signifies the converse (the components are multiplied).

Figure 1 relays a monotonic behavior for all four functions. A greater reduction in HbA1c or weight translates to added value for a given dula dose. The utility for weight depends on the change from baseline for HbA1c. If the change from baseline is $\geq x\%$ (actual value not shown for proprietary reasons), then the change will follow the red line. If the change is $< x\%$, then the change will follow the blue dashed line.

Increases in HR or DBP are considered undesirable. There is no increase in the utility for reductions in HR or DBP; however, neither HR nor DBP can exceed the value of 1. The utility penalizes for increases in either of these safety parameters; in fact, with significant increases in HR or DBP, the CUI decreases to 0 regardless of efficacy parameters (because of the multiplicative nature of the CUI). This is an advantage over the more commonly used additive CUI. With an additive CUI, efficacy can overwhelm safety; whereas, with the multiplicative CUI, a strong safety signal can trump efficacy.

To illustrate the CUI, pharmacodynamic (PD) models for the four end points previously developed and based on dula phase 1 data were applied. Although there is still uncertainty in these models, they are considered to be our best estimate of dula attributes given the available data and current knowledge about the mechanism of action for dula and, therefore, they offer an early assessment of the probability of trial success to support the business case. **Figure 2** illustrates the dose responses predicted from the PD models for the four measurements, and **Figure 3** shows application of the CUI based on these four responses. **Figure 2** shows four plots corresponding to the dose responses from the "most likely models" and resulting utilities for each component of the CUI. The top left plot shows that with increasing doses of dula, greater reductions in HbA1c are observed, and this is reflected as an increase in the utility function. Similarly, greater reductions in weight translate into an increasing utility function (top right plot). The bottom two plots show that with greater dula doses, the PD model predicts increases in HR and DBP, and the utilities for these two components of the CUI decrease with each dose until the value of 0 is reached.

Figure 3 shows the utility for each of the doses from the most likely model, illustrating how the four responses map to a single curve that yields the resulting CUI for each dula dose. The utilities (as a function of dose) go in

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

opposite directions for HbA1c and weight compared to HR and DBP. The lowest and highest dula doses have a CUI value below 1; however, each dose between the lowest and highest doses has a CUI above 1. These dula doses are predicted by the "most likely models" to be therapeutically optimal, reflecting the ideal balance of these four measures based on the CUI (see Simulation Process for further information on the "most likely models").

As mentioned previously, there is substantial uncertainty in these PD models. They are based on exposures of no more than 5 weeks, and they rely on assumptions and literature data to project 6 months for safety and 12 months

for efficacy. The decision rules require a sufficient level of evidence to be met (i.e., sufficient certainty) in order to select doses. After learning more about the effect of dula in this study about the effect of dula, the CUI for dula may turn out to be different from those suggested by these PD models.

Longitudinal Modeling
Longitudinal models are built to help understand the long-term effects of each treatment based on early observations of the four end points. For example, a model is built for the change in HbA1c through time for each treatment arm. As the study accrues data, this longitudinal
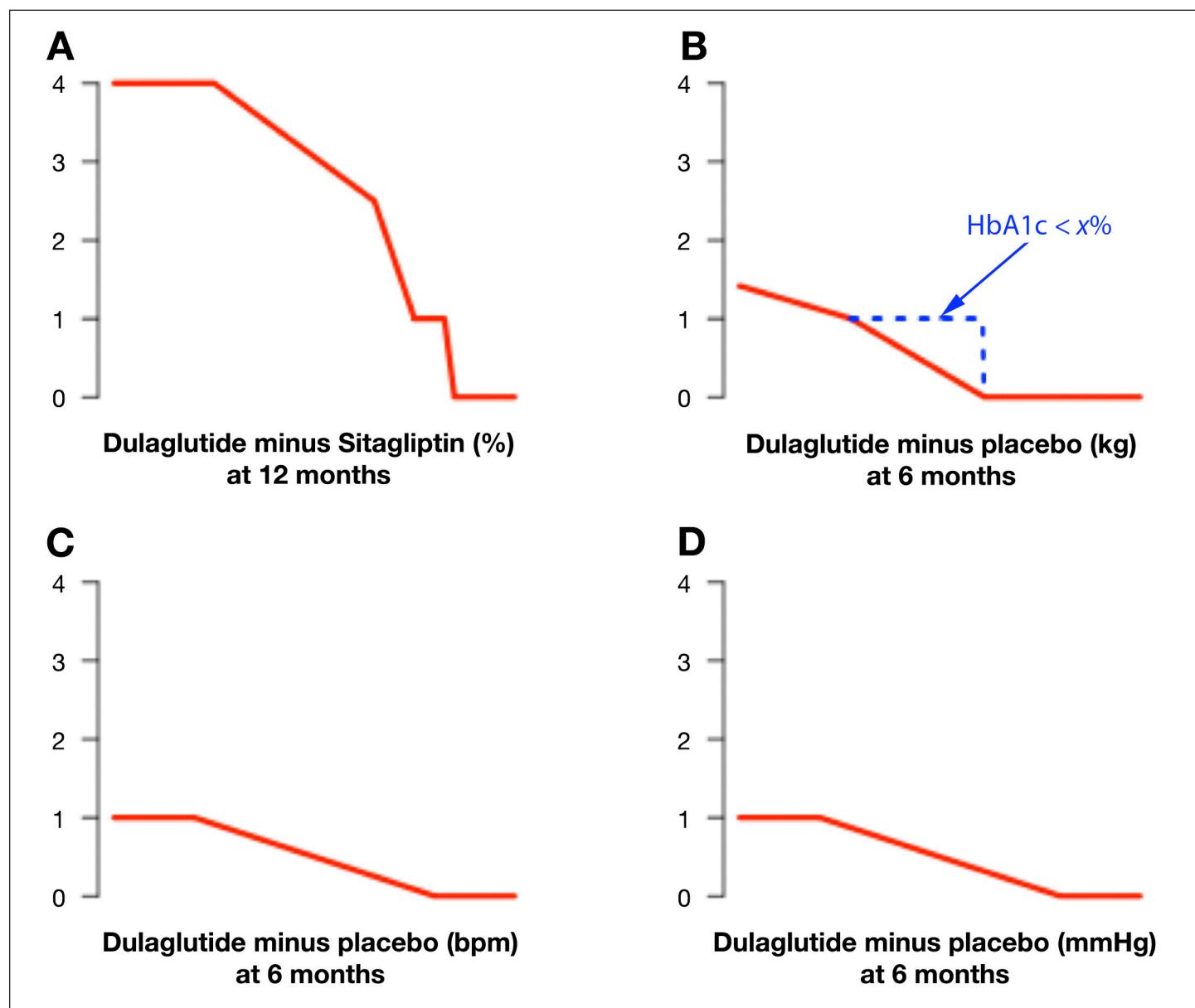

Figure 1. Utility components for the CUI plotted for **(A)** HbA1c, **(B)** weight, **(C)** HR, and **(D)** DBP. The utility components are a function change from baseline (CFB) relative to sitagliptin for HbA1c and placebo for the remaining measures. Note: *x*-axis not shown for proprietary reasons.

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

modeling creates a bridge between the early and later time periods and "learns" about the relative changes over time, providing information about the end points while the study is ongoing. The mean HbA1c at time t is modeled as a function of the final 12-month HbA1c change:

$$\exp(\gamma_t)\theta(d)$$

The γ_t parameters determine the mean at time t, relative to the 12-month end point, where $\theta(d)$ is the mean at 12 months for dose 'd'. A value of $\gamma_t = 0$ implies that the mean at time t is the same as the 12-month mean θ. The γ_t parameters are assumed to be identical across all doses, while the mean for placebo is assumed to be constant across time. This model was selected because it is flexible. This robust approach can be adapted to any growth scenario for each end point. Prior information on the longitudinal parameters can be easily incorporated. Separate versions of the longitudinal model were carried out for sitagliptin and for the dula doses. Placebo was assumed to have a constant mean HbA1c over time.
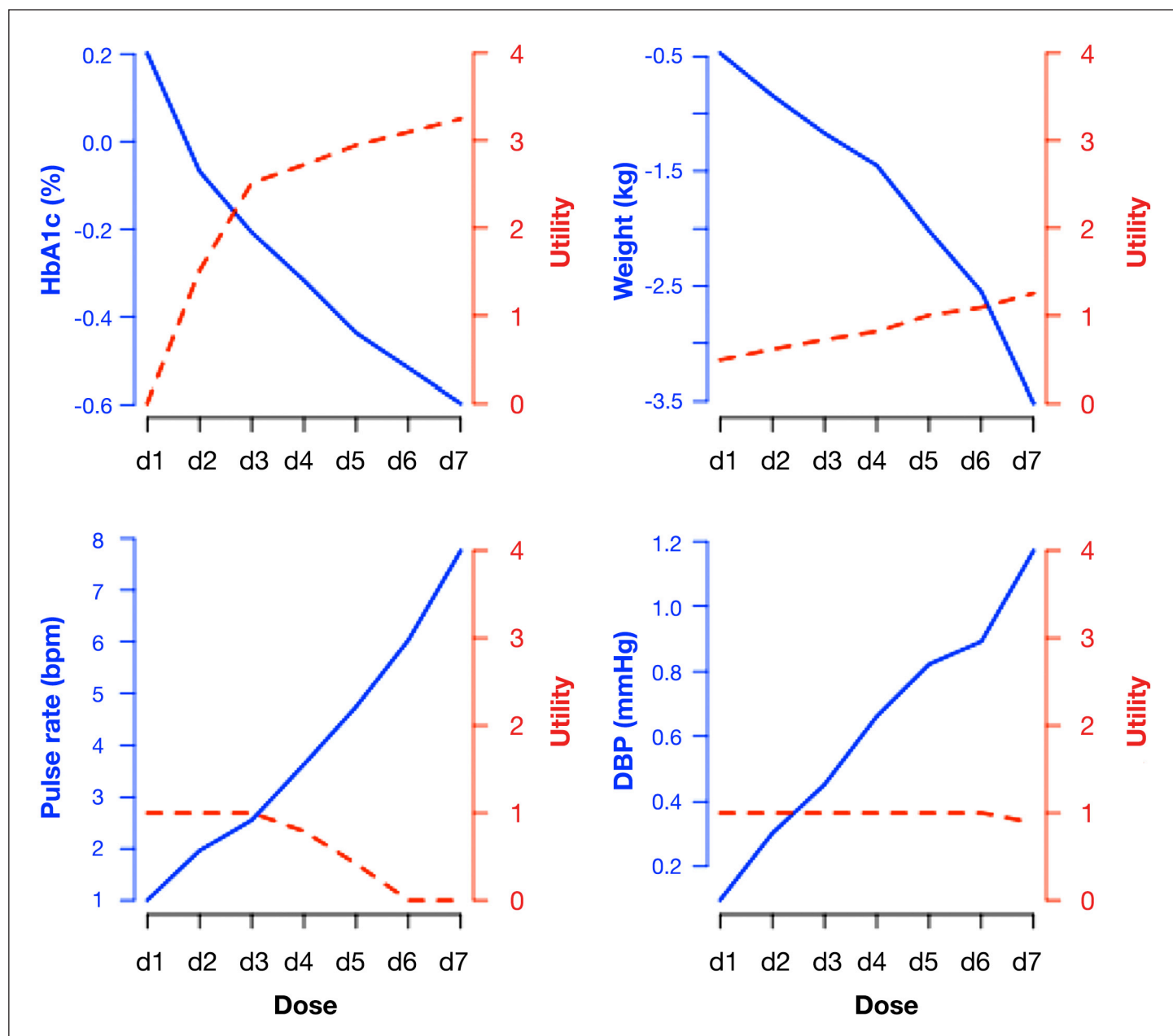

Figure 2. Change from baseline relative to comparator and corresponding values from utility components. Plot of the change from baseline of HbA1c, weight, HR, and DBP based on the most likely model and the corresponding utility component values.

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

This robust approach allows the relative means between time points to be determined by data, i.e., the longitudinal growth (decay) model. Prior information about the time course was explicitly incorporated to improve design efficiency. (See the next section on Bayesian Methods and Decision Rules for an explanation of priors). Prior distributions on γ_t are important during the adaptive steps in this trial because the amount of 12-month data may be limited or nonexistent. Independent normal distributions are selected as prior distributions for each of the γ_ts. Prior distributions are subjectively derived from prior data on this compound and on compounds in the same class.

Bayesian Methods and Decision Rules

Current knowledge about a parameter of interest, such as mean HbA1c for a given dose, is represented by the posterior probability distribution and is formed by updating prior distribution with data from the experiment. Prior probability distribution represents a probability distribution that transparently represents the belief of the clinical team based on expertise, which was informed by early-phase clinical results from dula and by data from other molecules in the same class. Posterior distribution builds on the prior, being guided by new data to create a current distribution of the phenomenon of interest. This Bayesian approach can be considered as a type of "active learning," where distribution is continually updated to represent the most up-to-date information.

Decision rules are based on thresholds for posterior probabilities for the CUI and thresholds for the predictive probabilities of meeting the primary objective of the study. The clinical utility of each dose is used to update the randomization probabilities and to assess decision rules. Safety and efficacy data drive the randomization scheme (by means of CUI) to allocate more patients to the most beneficial doses and fewer patients to less beneficial doses.

A major strength of the Bayesian approach is the ease with which predictions about future observations can be made. As part of the Bayesian approach, predictive probabilities are essential for designing clinical trials and have become a natural and useful tool for monitoring ongoing clinical trials.[5,6] The predictive probability can be described as a forward-thinking statistical tool that relates the probability of observing responses in future patients given current data. Based on cumulative information at the time, a predictive probability can be used to understand the probability of a positive result by the end of a trial. In this way, a predictive probability can be used
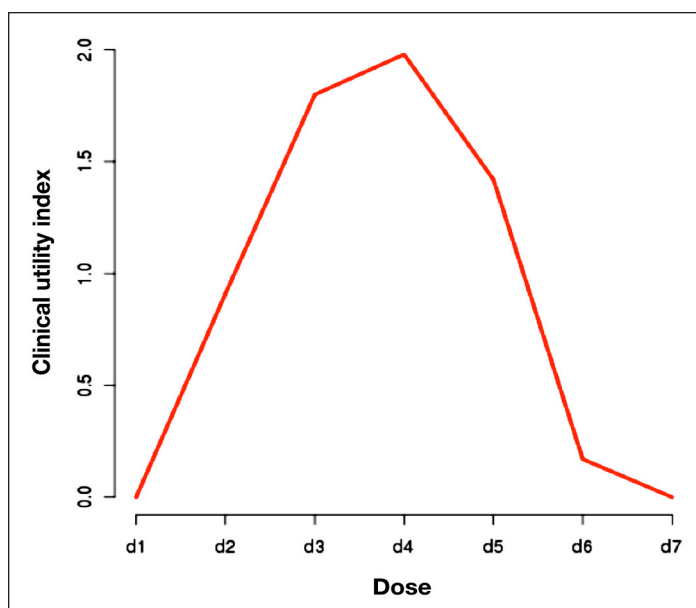

Figure 3. Plot of CUI derived by multiplying the four component utility measures. Plot of the CUI based on the most likely model for all 7 dulaglutide doses. The CUI for each of the dulaglutide doses from the "most likely" model is plotted in red.

to determine efficiently whether a trial is likely to be conclusive. Therefore, the predictive approach is used to evaluate whether a given sample size is appropriate for demonstrating the desired conclusion.

Based on data from stage 1, predictive probabilities provide useful information to determine the sample size for stage 2.[6] Bayesian approaches to determine sample size are discussed by Adcock[7] and Joseph and Belisle.[8] In the case of dula-dose selection, a predictive probability calculation will be used to choose between two sample-size schemes for stage 2 to ensure that a sufficient number of patients are enrolled to meet the study objectives. If the predictive probability of showing superiority to sitagliptin with a total of 263 patients per arm (including data on patients in the same arm from stage 1) was ≥85%, then 263 patients would be used; otherwise, 333 patients per arm would be used. In addition, stage 2 sample size will be augmented, if necessary, to ensure that at least 70% of the patients in each treatment arm come from stage 2 in order to mitigate against selection bias that may be introduced into the final analysis by including stage 1 data.

Final Analysis and Type I Error

The primary objective herein is to demonstrate that glycemic control of a high dose of dula is noninferior to that of sitagliptin at 12 months, as measured by change from baseline in HbA1c, with a noninferiority margin of 0.25%. This relatively strict margin was selected based on the previously reported variable efficacy of sitagliptin[9]

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

and limited efficacy data after 12 months of treatment[9] available during study design. This conservative margin will reduce the probability of falsely concluding non-inferiority. The final analysis (assuming that dose decision occurred) pools data from all randomized patients (the intent-to-treat population) from selected treatment arms (i.e., selected dula doses, placebo, and sitagliptin) and assesses the end point with a traditional analysis of covariance model, using the last observation with carried-forward imputation for missing HbA1c data. The prior distributions used in the Bayesian adaptive components of the design only affect the selection of doses for stage 2; they have no impact on the final frequentist-based analysis other than a (potentially) small increase in type I error, which will be adjusted for in the final analysis. Each of the selected dula doses will be tested for noninferiority and superiority to sitagliptin (at 12 months) and for superiority to placebo (at 6 months), resulting in 6 hypotheses if two doses are selected and 3 hypotheses if one dose is selected.

Because the entire trial is intended to be a phase-3 confirmatory trial, strong control of a type I error (i.e., the chance of observing a false-positive result) must be maintained, meaning that the overall probability of at least one hypothesis rejection being made falsely is less than a prespecified type I error level. Because potentially six hypotheses will be tested, inflation of a type I error caused by multiple comparisons may be an issue. To address this, a tree-gatekeeping strategy was chosen. Strong control of a type I error is guaranteed with the tree-gatekeeping strategy when applied to a fixed design.[10]

In this inferentially seamless design, selection bias could inflate a type I error,[11] the probability of a false positive. For example, a false-positive finding in stage 1 could lead to a false-positive conclusion in the overall analysis because the overall analysis combines data from patients in stages 1 and 2. To mitigate selection bias, ≥70% of patients within each treatment arm are enrolled during stage 2, and a conservative nominal α-level of 0.02 will be used in the final analysis to ensure that the overall type I error is maintained at 0.025. Simulations demonstrated that this adjustment will be more than adequate to ensure strong control of a type I error. In addition, the study is designed to have sufficient power to assess primary objectives with stage 2, alone, if there is a concern about selection bias.

Simulation Process

Modeling and simulation were critical to the development of this design. As discussed in Geiger and coauthors,[1]

dose–response models for HbA1c and weight and exposure-response models for HR and DBP were created. Seven dula doses, placed incrementally over a 12-fold dose range, were chosen to allow for adequate exploration of the dose-response curves. Thousands of trials over a wide range of scenarios were simulated, with varying assumptions about response measures (dose response and longitudinal profile), enrollment rates, and dropout rates. Typically, 1000 trial replicates per scenario were simulated with the exception of null-set simulations, which required 10,000 replicates. Each of the simulations was initialized by its own input file consisting of parameters for the particular simulation and a random initialization seed. These independent jobs were run in parallel on a Sun Grid Engine cluster; the simulation program was written in Fortran 77.

Creation of the adaptive algorithm and finalization of the decision rules were iterative processes. Emphasis was placed on the selection accuracy of particular dula doses ("right" doses) and how often the trial continued into stage 2 with dula doses that met the prespecified dose-selection criteria. Likewise, probabilities of early termination, randomization of patients to ineffective treatment arms, and incorrect termination decisions were carefully accessed by simulation.

A wide range of scenarios were simulated, and the algorithm was modified iteratively based on its performance. We analyzed data from dula preclinical studies, dula phase 1 studies, and published data on other GLP-1 analogs to understand the dula drug-disease state and to synthesize the information into PD models for end points of interest. These models can be referred to as the "most likely models" because they represent our best understanding of the dula-dose response (shown in **Figure 4**).

Results

Pharmacodynamic modeling predicted that several dula doses would meet the predefined 88% safety and efficacy criteria for dose selection. Using the proposed adaptive design, trials simulated from the "most likely models" correctly identified at least one dose that met these criteria. In contrast, simulation studies for a traditional fixed-dose design (assuming 50 patients assigned to placebo and four dose groups consisting of dula 0.5, 1, 2, and 3 mg) indicated a low chance of success (6 to 12% at 12 weeks or 26 weeks, respectively) to yield adequate information for a dose decision due to the failure to identify a dose that satisfied strict criteria for safety

Application of Adaptive Design Methodology in Development of a Long acting Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

measures. In a fixed design, the variability in the safety end points was too great to assess with adequate precision.

Figure 4 shows a plot of results from trial simulations using PD models and illustrates why adaptive randomization performs better than a fixed design. The probability of dose selection follows the pattern of the CUI. The sample size also follows the value of the CUI; more patients are allocated to safe and effective doses compared with doses that are less beneficial because the adaptive design considers the available data to learn about dose response (compared with a fixed design in which the number of patients is equally allocated to each dose regardless of their observed safety and efficacy).

Not only did the algorithm select doses when it should, it also stops the study when it should, as illustrated in **Table 1**, with six different response scenarios in which the annual dropout rates vary between 0 and 20% and the enrollment rates between five and eight patients per week (see **Appendix A** for longitudinal dose-response models). The results from the first and third scenarios show that despite (simulated) robust efficacy, the algorithm stops the trial before stage 2 at least 90% of the time when the mean increase in HR is +10 beats per minute (bpm) compared with placebo or when the mean increase in DBP is +5 mm Hg compared with placebo. When either HR *or* DBP is mildly elevated (+5 bpm and +2 mm Hg, respectively) and there is borderline efficacy, the algorithm stops the trial without entering into stage 2 more than 50% of the time. When both HR *and* DBP are mildly elevated, the algorithm stops the trial more than 85% of the time even with borderline efficacy. A weight gain of 5 kg causes the algorithm to stop the trial 100% of the time.

To evaluate a type I error, scenarios were assessed in which dula lacked efficacy (inferior to sitagliptin). In these scenarios, the type I error was well controlled. In simulations with no dropouts, all type I error estimates were at least 2 simulation standard errors below the targeted 5% (2-sided). In simulations with dropouts, the type I error was controlled to the same level of an analogous fixed-dose design.

The simulations provided robust evidence that the type I error is controlled. In fact, the simulations indicate that the nominal level is conservative, and the resulting type I error is <5% (2-sided). The decision rules (based on the CUI and the predictive probability of noninferiority) affect the type I error by limiting the opportunities to enter into stage 2 to only doses that show evidence of
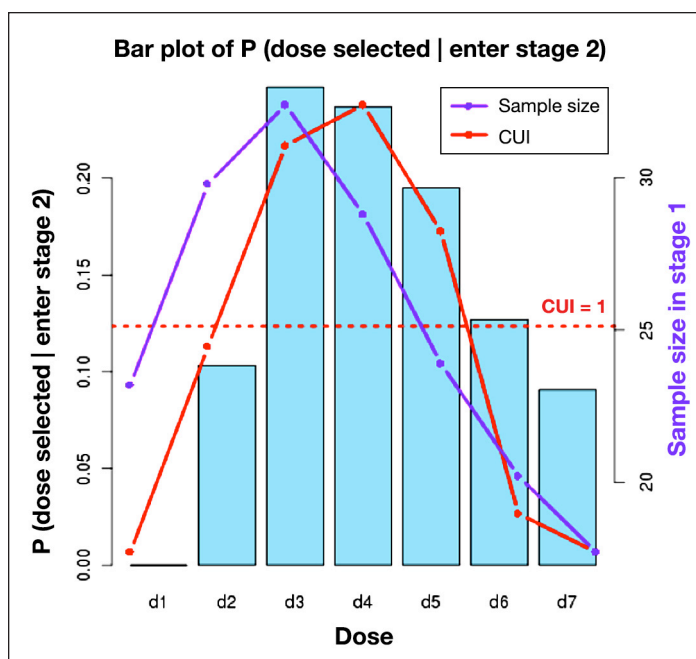
Figure 4. Plot of operating characteristics of the adaptive algorithm for the most likely model. A bar plot of the P (dose is selected given that stage 2 was conducted) is given with the scale on the left *y*-axis. The purple line plots the sample size with the corresponding scale given on the right *y*-axis. The CUI is plotted in red with no scale given. A reference line for CUI = 1 is provided.

being both safe and efficacious. The probability of making a false decision in the final analysis (based on the selected doses studied in both stage 1 and stage 2) is greatly reduced when there are safety issues because the study tends to end at stage 1 because of the "futility" rules in this study. The scenarios studied to assess the type I error were conservative, with the goal of maximizing the type I error, even if the scenario was unrealistic. In all of the scenarios, there were no safety issues, and the dula doses were exactly noninferior; nonetheless, the proportion of trials with a false positive was no worse than what one would expect from a fixed design.

Simulations demonstrated that an enrollment rate greater than eight patients per week resulted in suboptimal design performance, i.e., a decreased probability that the algorithm would select the "right" dose(s) or would stop the study for futility when it should. As a result, a mean enrollment rate of no more than eight patients per week was targeted.[12] Other factors to consider are frequency of updating the randomization scheme and assessing the decision rules. Improvement in operating characteristics should be weighed against the logistical implications in a clinical trial setting (drug supply, data management, statistical analysis, and DMC reviews).[12] The ideal frequency of updates depends in part on the time-to-effect of the biomarkers being studied, which are related

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

to how much learning can take place within a given interval of time between updates. In this study, through simulations, updating every 2 weeks and assessing decision rules after the enrollment of 200 patients were found to be adequate.

This study design was more efficient with a controlled enrollment rate of no more than eight patients per week to allow the adaptive algorithm to "learn" and to adjust the randomization properties appropriately.

Discussion

Other adaptive designs, such as group sequential designs, may also be considered for diabetes studies. Group sequential designs employ early stopping rules for either the presence of or lack of efficacy.[13-16] Early stopping

for efficacy is typically not accepted because regulators require a minimum amount of exposure to a given therapy for T2DM. Early-stopping rules for lack of efficacy could be applied, but this fails to incorporate information about safety in the decision to terminate the study. Often the question in developing a diabetes drug is not whether or not the drug is efficacious but, rather, if there is a therapeutic window in which the drug is both efficacious and safe. For this purpose, safety data need to be explicitly incorporated in clinical trial adaptations, which group sequential designs fail to do. Group sequential designs also use a fixed randomization scheme, allocating patients to all treatments with fixed probabilities. In contrast, AWARD-5 uses an adaptive randomization scheme, randomizing patients to doses in proportion to the desirability of the doses. This is a more efficient randomization scheme, allocating patients

Table 1. Simulation Results for Scenarios Assessing the "Futility" Rules for Safety.					
Simulation Scenario	Execution Parameters		Decision-Point Metrics		
	DO	ER	P(GO)	P(CAP)	P(FUT)
HR elevated + 10 bpm, great efficacy, all doses HbA1c model = 1, Weight loss model = 1, HR model = 3, DBP model = 3	0	5	0.007	0.007	0.986
	0	8	0.012	0.013	0.975
	20	5	0.006	0.003	0.991
	20	8	0.008	0.005	0.987
HR elevated + 5 bpm, borderline efficacy, all doses HbA1c model = 3, Weight loss model = 5 HR model = 5, DBP model = 3	0	5	0.318	0.348	0.334
	0	8	0.260	0.387	0.353
	20	5	0.290	0.355	0.355
	20	8	0.255	0.377	0.368
DBP elevated + 5 mmHg, great efficacy, all doses HbA1c model = 1, Weight loss model = 1 HR model = 4, DBP model = 2	0	5	0.059	0.085	0.856
	0	8	0.079	0.121	0.800
	20	5	0.068	0.096	0.836
	20	8	0.093	0.114	0.793
DBP elevated + 2 mmHg, borderline efficacy, all doses HbA1c model = 3, Weight loss model = 5 HR model = 4, DBP model = 4	0	5	0.388	0.373	0.239
	0	8	0.350	0.398	0.252
	20	5	0.372	0.402	0.226
	20	8	0.336	0.423	0.241
HR and DBP elevated with borderline efficacy, all doses HbA1c model = 3, Weight loss model = 5 HR model = 5, DBP model = 4	0	5	0.099	0.249	0.652
	0	8	0.078	0.263	0.659
	20	5	0.090	0.245	0.665
	20	8	0.089	0.275	0.636
Weight gain of +5 kg, all doses HbA1c model = 1, Weight loss model = 3 HR model = 4, DBP model = 3	0	5	0.000	0.000	1.000
	0	8	0.000	0.000	1.000
	20	5	0.000	0.000	1.000
	20	8	0.000	0.000	1.000
DO, dropout rate (%); ER, enrollment rate (patients per week); FUT, futility; N1, noninferiority; P(CAP), probability of stopping for reaching the cap; P(FUT), probability of stopping in stage 1 for futility; P(GO), probability of continuing the study into stage 2; Sup, superiority.					

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

to desirable doses instead of treating all doses equally despite data indicating that some doses are less desirable than others.

Our study design owes much to the Acute Stroke Therapy by Inhibition of Neutrophils (ASTIN) trial design[3,17,18] but has incorporated several additional novel features. Adaptations in ASTIN were based on a single efficacy measure, whereas our adaptive algorithm and dose-selection criteria are based on a CUI inclusive of safety, desirability, and efficacy measures. The ASTIN trial was intended to be seamless in design, but the trial never progressed to that stage because it was terminated due to futility. Our design uses a Bayesian decision theoretical approach; the ASTIN trial[17] relied solely on an assessment of the posterior probability of clinically meaningful effects to make decisions. In addition, the ASTIN trial suffered from unrealistic assumptions and limited updating of the linear regression models used in its longitudinal modeling. Our trial used exponential growth models, which are not susceptible to the parameterization issues described by Berry and coauthors[3] and by Grieve and coauthors[18] and are much more robust to influential observations that may not be representative of the population. We update the longitudinal models from the start of the trial.

Conclusions

Patients may be exposed to dula for up to 12 months in stage 1, potentially enabling detection of safety signals much earlier (in terms of the drug development of diabetes compounds) than is possible in a development program that uses a shorter, fixed phase 2 study design. Additionally, this approach enables more safety data to be available at the time of dose decision, which increases the likelihood of choosing the best dose(s) for continued evaluation in phase 3 studies. The adaptive design features of this trial offer a safer and more effective approach for the evaluation of dula than a fixed design, given the enhanced probability of correctly identifying a dose and its additional safety features.

References:

1. Geiger MJ, Skrivanek Z, Gaydos B, Chien J, Berry S, Berry D, Anderson JH, Jr. An adaptive, dose-finding, seamless phase 2/3 study of a long-acting glucagon-like peptide-1 analog (dulaglutide): trial design and baseline characteristics. J Diab Sci Tech. 2012;6(6):1319–27.

2. West M, Harrison J. Bayesian forecasting and dynamic models. 2nd ed. New York: Springer-Verlag; 1997.

3. Berry DA, Müller P, Grieve AP, Smith M, Parke T, Blazek R, Mitchard R, Krams M. Adaptive bayesian designs for dose-ranging drug trials. In: Gatsonis C, Kass RE, Carlin B, Carriquiry A, Gelman A, Verdinelli I, West M, editors. Case studies in bayesian statistics V. New York: Springer-Verlag; 2001;99–181.

4. Korsan B, Dykstra K, Pullman WM. Transparent trade-offs: a clinical utility index (CUI) openly evaluates a product's attributes and chance of success. Pharmaceutical Executive. 2005;[cited 2011]. Available from: *http://www.pharsight.com/library/PE5-58-05e.pdf*.

5. Berry DA. Bayesian statistics and the efficiency and ethics of clinical trials. Stat Sci. 2004;19(1):175–87.

6. Lecoutre B, Dzerko G, Grouin JM. Bayesian predictive approach for inference about proportions. Stat Med. 1995;14(9–10):1057–63.

7. Adcock CJ. Sample size determination: a review. Statistician. 1997;46(2):261–83.

8. Joseph L, Belisle P. Bayesian sample size determination for normal means and differences between normal means. Statistician. 1997;46(2):209–26.

9. Januvia [package insert]. Whitehouse Station, NJ: Merck & Co., Inc.; 2006.

10. Dmitrienko A, Tamhane AC, Wiens BL. General multistage gatekeeping procedures. Biom J. 2008;50(5):667-77. Northwestern University, Evanston, IL: Department of Industrial Engineering and Management Sciences 2007; Working Paper No. 07-06. Available from: *http://www.iems.northwestern.edu/research/papers.html*. Accessed September 10, 2012.

11. Zelen M. Play-the-winner rule and the controlled clinical trial. J Am Stat Assoc. 1969;64(325):131–46.

12. Spencer K, Colvin K, Braunecker B, Brackman M, Ripley J, Hines P, Skrivanek Z, Gaydos B, Geiger MJ. Operational challenges and solutions with implementation of an adaptive, seamless phase 2/3 study. J Diab Sci Tech. 2012;6(6):1296–1304.

13. Pocock SJ. Group sequential methods in the design and analysis of clinical trials. Biometrika. 1977;64(2):191–9.

14. O'Brien PC, Fleming TR. A multiple testing procedure for clinical trials. Biometrics. 1979;35(3):549–56.

15. Armitage P. Sequential medical trials. 2nd ed. New York, NY: John Wiley and Sons, Inc; 1975.

16. DeMets DL, Ware JH. Group sequential methods for clinical trials with a one-sided hypothesis. Biometrika. 1980;67(3): 651–60.

17. Krams M, Lees KR, Hacke W, Grieve AP, Orgogozo JM, Ford GA. Acute stroke therapy by inhibition of neutrophils (ASTIN): an adaptive dose-response study of UK-279,276 in acute ischemic stroke. Stroke. 2003;34(11):2543–8.

18. Grieve AP, Krams M. ASTIN: a Bayesian adaptive dose-response trial in acute stroke. Clin Trials. 2005;2(4):340–51.

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

Appendix A. HbA1c Longitudinal Dose–Response Models (1–11)

Table 1. HbA1c Model 1									
WK	PLA	0.25 mg LY	0.50 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	-0.30	-0.37	-0.49	-0.78	-0.93	-1.11	-1.18	-0.49
4	0	-0.36	-0.44	-0.59	-0.93	-1.12	-1.33	-1.42	-0.59
8	0	-0.40	-0.49	-0.65	-1.04	-1.24	-1.48	-1.58	-0.65
12	0	-0.43	-0.52	-0.70	-1.10	-1.32	-1.57	-1.68	-0.70
26	0	-0.47	-0.58	-0.77	-1.22	-1.47	-1.75	-1.86	-0.77
39	0	-0.49	-0.59	-0.79	-1.26	-1.50	-1.79	-1.91	-0.79
52	0	-0.49	-0.60	-0.80	-1.27	-1.52	-1.81	-1.93	-0.80
LY, LY2189265; PLA, placebo; WK, week.									

Table 2. HbA1c Model 3									
WK	PLA	0.25 mg LY	0.50 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	-0.49	-0.49	-0.49	-0.49	-0.49	-0.49	-0.49	-0.49
4	0	-0.59	-0.59	-0.59	-0.59	-0.59	-0.59	-0.59	-0.59
8	0	-0.65	-0.65	-0.65	-0.65	-0.65	-0.65	-0.65	-0.65
12	0	-0.70	-0.70	-0.70	-0.70	-0.70	-0.70	-0.70	-0.70
26	0	-0.77	-0.77	-0.77	-0.77	-0.77	-0.77	-0.77	-0.77
39	0	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79	-0.79
52	0	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80	-0.80
LY, LY2189265; PLA, placebo; WK, week.									

Table 3. Weight Loss Model 1									
WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	-0.2	-0.4	-0.6	-0.7	-0.8	-0.8	-1.1	0
4	0	-0.4	-0.8	-1.1	-1.3	-1.4	-1.4	-1.7	0
8	0	-0.6	-1.4	-1.9	-2.3	-2.4	-2.6	-2.9	0
12	0	-0.8	-1.8	-2.5	-2.9	-3.1	-3.4	-3.7	0
26	0	-1.1	-2.4	-3.4	-4	-4.3	-4.5	-4.8	0
39	0	-1.2	-2.6	-3.7	-4.4	-4.7	-5	-5.3	0
52	0	-1.2	-2.7	-3.8	-4.5	-4.8	-5.2	-5.5	0
LY, LY2189265; PLA, placebo; WK, week.									

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

Appendix A. HbA1c Longitudinal Dose–Response Models (1–11)

Table 4.
Weight Loss Model 3

WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	5	5	5	5	5	5	5	0
4	0	5	5	5	5	5	5	5	0
8	0	5	5	5	5	5	5	5	0
12	0	5	5	5	5	5	5	5	0
26	0	5	5	5	5	5	5	5	0
39	0	5	5	5	5	5	5	5	0
52	0	5	5	5	5	5	5	5	0

LY, LY2189265; PLA, placebo; WK, week.

Table 5.
Weight Loss Model 5

WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0.00	-0.31	-0.31	-0.31	-0.31	-0.31	-0.31	-0.31	1.41
4	0.00	-0.85	-0.85	-0.85	-0.85	-0.85	-0.85	-0.85	2.14
8	0.00	-1.57	-1.57	-1.57	-1.57	-1.57	-1.57	-1.57	3.14
12	0.00	-2.00	-2.00	-2.00	-2.00	-2.00	-2.00	-2.00	3.72
26	0.00	-2.50	-2.50	-2.50	-2.50	-2.50	-2.50	-2.50	4.41
39	0.00	-2.56	-2.56	-2.56	-2.56	-2.56	-2.56	-2.56	4.51
52	0.00	-2.59	-2.59	-2.59	-2.59	-2.59	-2.59	-2.59	4.54

LY, LY2189265; PLA, placebo; WK, week.

Table 6.
Heart Rate Model 3

WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	5	5	5	5	5	5	5	0
4	0	10	10	10	10	10	10	10	0
8	0	10	10	10	10	10	10	10	0
12	0	10	10	10	10	10	10	10	0
26	0	10	10	10	10	10	10	10	0
39	0	10	10	10	10	10	10	10	0
52	0	10	10	10	10	10	10	10	0

LY, LY2189265; PLA, placebo; WK, week.

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

Appendix A. HbA1c Longitudinal Dose–Response Models (1–11)

Table 7. Heart Rate Model 4									
WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0
LY, LY2189265; PLA, placebo; WK, week.									

Table 8. Heart Rate Model 5									
WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	5	5	5	5	5	5	5	0
4	0	5	5	5	5	5	5	5	0
8	0	5	5	5	5	5	5	5	0
12	0	5	5	5	5	5	5	5	0
26	0	5	5	5	5	5	5	5	0
39	0	5	5	5	5	5	5	5	0
52	0	5	5	5	5	5	5	5	0
LY, LY2189265; PLA, placebo; WK, week.									

Table 9. Diastolic Blood Pressure Model 2									
WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	2	2	2	2	2	2	2	0
4	0	5	5	5	5	5	5	5	0
8	0	5	5	5	5	5	5	5	0
12	0	5	5	5	5	5	5	5	0
26	0	5	5	5	5	5	5	5	0
39	0	5	5	5	5	5	5	5	0
52	0	5	5	5	5	5	5	5	0
LY, LY2189265; PLA, placebo; WK, week.									

Application of Adaptive Design Methodology in Development of a Long acting
Glucagon-like Peptide-1 Analog (Dulaglutide): Statistical Design and Simulations

Skrivanek

Appendix A. HbA1c Longitudinal Dose–Response Models (1–11)

Table 10. Diastolic Blood Pressure Model 3									
WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0	0	0
LY, LY2189265; PLA, placebo; WK, week.									

Table 11. Diastolic Blood Pressure Model 4									
WK	PLA	0.25 mg LY	0.5 mg LY	0.75 mg LY	1 mg LY	1.5 mg LY	2 mg LY	3 mg LY	Sitagliptin
2	0	2	2	2	2	2	2	2	0
4	0	2	2	2	2	2	2	2	0
8	0	2	2	2	2	2	2	2	0
12	0	2	2	2	2	2	2	2	0
26	0	2	2	2	2	2	2	2	0
39	0	2	2	2	2	2	2	2	0
52	0	2	2	2	2	2	2	2	0
LY, LY2189265; PLA, placebo; WK, week.									