

Application Type	BLA
STN	125546/0
CBER Received Date	July 24, 2014
PDUFA Goal Date	March 24, 2015
Division / Office	DVRPA /OVRR
Committee Chair	Margaret C. Bash, M.D.
Clinical Reviewer(s)	Anuja Rastogi, M.D.
Project Manager	Edward Wolfgang, MSA; Kirk Prutzman, Ph.D.; Ramachandra Naik, Ph.D.
Priority Review	Yes
Reviewer Name(s)	Tsai-Lien Lin, Ph.D. Lead Mathematical Statistician
Review Completion Date / Stamped Date	
Supervisory Concurrence	A. Dale Horne, Dr. P.H. Chief, Vaccine Evaluation Branch, DB, OBE
	Estelle Russek-Cohen, Ph.D. Director, Division of Biostatistics, OBE
Applicant	Novartis Vaccines and Diagnostics, Inc.
Established Name	Meningococcal Group B Vaccine, rMenB+OMV NZ
(Proposed) Trade Name	Bexsero®
Pharmacologic Class	Vaccine
Formulation(s), including Adjuvants, etc	Each dose contains 50µg of each of the three purified recombinant protein antigens NadA, NHBA and fHbp, with 25µg of OMV measured as amount of total protein containing the PorA P1.4, and 1.5 mg of aluminum hydroxide per 0.5 ml dose
Dosage Form(s) and Route(s) of Administration	0.5 mL suspension for intramuscular injection as a single dose pre-filled syringe
Dosing Regimen	Two doses (0.5mL each) with an interval of at least 1 month between doses
Indication(s) and Intended Population(s)	Active immunization against invasive disease caused by <i>Neisseria meningitidis</i> serogroup B strains in individuals from 10 through 25 years of age

Table of Contents

1. Executive Summary	3
2. Regulatory Background	5
3. Sources of Data and Other Information Considered in the Review	5
3.1 REVIEW STRATEGY	5
3.2 BLA/IND DOCUMENTS THAT SERVE AS THE BASIS FOR THE STATISTICAL REVIEW	5
4. Discussion of Individual Assays.....	7
4.1 (B)(4) POTENCY TEST	7
4.2 (B)(4) ASSAYS	13
4.3 ENDOTOXIN SPECIFICATION	21
5. Conclusions and Recommendations	23

1. EXECUTIVE SUMMARY

Novartis submitted BLA 125546/0 to seek licensure of the multicomponent Meningococcal group B vaccine, Bexsero®, indicated for active immunization against invasive disease caused by *Neisseria meningitidis* serogroup B strains of individuals from 10 through 25 years of age. This review focuses on the (b)(4) immunogenicity potency test, the (b)(4) (which were used in the (b)(4) immunogenicity potency test and considered part of the potency test), and the endotoxin specification. This vaccine has been licensed in the EU, Canada, and Chile, but some concerns over several of the assays still remain.

(b)(4) Immunogenicity Potency Test

The original potency assay, used since before EU approval in 2013, has very large variability. The current immunogenicity potency test differs from the original test procedure in the assay format (using only (b)(4) dilutions instead of (b)(4) per group instead of (b)(4), and --(b)(4)-- assays to obtain the final reportable value instead of (b)(4) assay) and the analysis model ----- (b)(4) ----- instead of ----- (b)(4) ----- . The system suitability criteria were determined by simulation, and the validation was conducted by using the 2012 validation data for the original potency test to generate all possible potency results under the (b)(4) model. In addition, dilutional linearity and accuracy were not assessed. The variability, although improved, appears to be still quite large based on either the simulated validation or the release data available. Because of the large variability, setting the potency specification is a challenge.

The (b)(4) assays, which are used to ----- (b)(4) ----- in the potency test, generally do not have an adequate control strategy in place. The system suitability criteria appear to be less than optimal and the sample validity criteria are lacking. The validation parameters of precision, linearity, and LLOQ do not have the same meanings in comparison to how they are usually defined. Therefore, it is not clear whether the (b)(4) assays could have contributed to the large variability of the immunogenicity potency test.

With high variability, the current immunogenicity potency test may only be able to detect substantial differences in potency and therefore is recommended to be used only as an interim test until further improvements can be implemented. The current specifications are likely to be adequate to ensure that no substantially subpotent lots are released. Further revision of potency specifications will be needed after the potency assay is improved. Novartis, in response to the 11/12/14 Information Request (IR), has committed to improve the immunogenicity test and the (b)(4) assays and to revise the potency specifications (Amendment 30, dated 12/1/14).

Endotoxin Specification

The concern of the current endotoxin upper limit --- (b)(4) ----- being too high was initially complicated with the uncertainty about the assay variability due to the huge variability observed in the stability data. The assay variability issue was satisfactorily

resolved, from the statistical perspective, after Novartis explained that only the stability lots initiated after October 2012 are tested by the modified method, which is consistent with the more consistent and precise stability data observed in the later stability lots. The variability of the current modified (b)(4) assay is found to be consistently (b)(4) in the datasets from the validation, the clinical lots, and available commercial lots and stability data tested after the implementation of the new improved assay. The modified assay, although having much smaller variability, generates much higher values compared to the previous assay. The data for the clinical lots also showed that the variability contributed by the manufacturing process is approximately the same magnitude as the assay variability.

Based on the clinical lots data used to set the specification and the (b)(4) production lots, this upper limit does reflect the current assay and process capability. The statistical reviewer defers to the product reviewers on the acceptability of this one-sided endotoxin upper limit.

2. REGULATORY BACKGROUND

Novartis has used the -----(b)(4)----- assay as the potency test for final drug product since before EU approval in 2013. To improve the assay performance, Novartis has made several changes over time to the test methods. The SOPs and the validation reports of the immunogenicity test and the (b)(4) methods, which measure the antibody levels of -----(b)(4)----- test, were submitted to IND (b)(4) Amendment 212 on 3/14/2014 and to BLA 125546.0.1 on 7/9/2014. CBER's information requests and comments regarding the (b)(4) potency assay and the (b)(4) assays were sent to Novartis on 4/16/2014. Novartis's responses to these CBER comments sent under IND (b)(4) were submitted to BLA 125546.0.4 on 7/29/14. Several additional IRs were also communicated to Novartis during the BLA review.

Concern about the endotoxin specification being too high was raised in the 5/27/14 pre-BLA meeting. Novartis provided justification for endotoxin specifications in BLA 125546.0.1 and Amendments 9 and 15. Three additional technical reports completed during the endotoxin assay development/improvement were submitted in Amendment 35 to address the questions raised by the review team.

3. SOURCES OF DATA AND OTHER INFORMATION CONSIDERED IN THE REVIEW

3.1 REVIEW STRATEGY

This review focuses on the (b)(4) immunogenicity potency test and the (b)(4)- assays used to test the sera from the ----(b)(4)----- potency test, about which CBER did not receive any information until the pre-BLA meeting. The (b)(4) assays are considered part of the immunogenicity potency test. However, for clarity, the immunogenicity potency test and the (b)(4) assays are discussed in sections 4.1 and 4.2 separately. Statistical review of the endotoxin specification was also conducted at the product reviewer's request and is discussed in section 4.3.

3.2 BLA/IND DOCUMENTS THAT SERVE AS THE BASIS FOR THE STATISTICAL REVIEW

- BLA 125546.0.1 dated 7/9/2014
 - Module 2.3 Quality Overall Summary
 - 2.3.S. Drug Substance
 - 2.3.P. Drug Product
 - Module 3. Quality
 - 3.2.P. Drug Product
 - 3.2.P.5.1. Specifications
 - 3.2.P.5.2. Analytical Procedure -----(b)(4)-----
Analytical Procedure -----(b)(4)-----
Analytical Procedure -----(b)(4)-----
 - 3.2.P.5.3. Validation of Analytical Procedures

Validation of Analytical Procedure-Intro

---(b)(4)-----

---(b)(4)----

endotoxin

----- (b)(4)-----

3.2.P.5.6. Justification of Specifications

- BLA 125546.0.4 dated 7/29/2014
Module 1.11.1. Quality Information Amendment ((b)(4) potency)
- BLA 125546.0.7 dated 8/19/2014
Module 3.2.P.5. 3. Validation of Analytical Procedures ((b)(4) potency)
- BLA 125546.0.9 dated 9/3/2014
Module 1.11.1. Quality Information Amendment (endotoxin)
- BLA 125546.0.11 dated 9/12/2014
Module 3.2.P.5.3. Validation of Analytical Procedures ((b)(4) potency)
- BLA 125546.0.12 dated 9/16/2014
Module 1.11.1 Quality Information Amendment –Potency Assay Fup
- BLA 125546.0.15 dated 9/29/2014
Module 1.11.1 Quality Information Amendment – IR Endotoxin Fup
- BLA 125546.0.16 dated 10/2/2014
Module 1.11.1 Quality Information Amendment – IR Potency Fup2
Module 3.2.P.5.4 Batch Analyses
- BLA 125546.0.30 dated 12/1/2014
Module 1.11.1 Quality Information Amendment – (b)(4) Potency Assay
- BLA 125546.0.35 dated 12/15/2014
Module 1.11.1 Quality Information Amendment – Endotoxin Fup2
Module 3.2.P.5.3 Validation of Analytical Procedures
Technical Report 292507
Technical Report 292581
Technical Report 294387
- IND (b)(4)-----

4. DISCUSSION OF INDIVIDUAL ASSAYS

4.1 (B)(4) POTENCY TEST

Description of (b)(4) Potency (b)(4) Assay

The (b)(4) potency assay is designed as a -----(b)(4)----- assay in ---(b)(4)--. The current assay format uses (b)(4) independent assays to determine the potency for each test sample. In each assay, -----
----- (b)(4)-----

----- The antibody concentrations against three recombinant proteins (287-953, 936-741, and 961c) and OMV for each individual serum sample are measured by (b)(4).

The antibody concentrations obtained are used to calculate the Relative Potency (RP) of the test vaccine against the reference vaccine by applying the -----(b)(4)-----
----- mathematical model as follows:

- -----(b)(4)----- models for -----(b)(4)----- are evaluated and assessed for mathematical "fit"
- For -----(b)(4)----- the following acceptance criteria are applied:
 - For system suitability (on the model for the reference lot alone)
 - Slope p-value (b)(4);
 - Sum of squares of Non-linearity \leq limits established (sometimes this is called a lack of fit sum of squares)
 - For sample suitability (on the full model of reference and sample lots)
 - Slope p-value (b)(4);
 - Sum of squares of Non-linearity \leq limits established
 - Sum of squares of Non-parallelism \leq limits established
- In case the assay is valid with ---(b)(4)-----, the results come from the ---(b)(4)----- with the greater common slope. In case the assay is valid with only ---(b)(4)----- range, the results come from the ---(b)(4)----- where the assay is valid. In case the assay is invalid with -----(b)(4)-----, the assay is invalid.

The valid assays are then used to obtain the reportable RP values:

- In case the assays from (b)(4) immunizations are valid, -----

----- (b)(4) -----

- In case only (b)(4) is valid, RP is calculated from the valid assay.

- In case the assays from (b)(4) immunizations are invalid, the (b)(4) of independent assays is invalid.

The system suitability criteria were based on simulated data due to limited historical (b)(4) data available. Assays of the reference standard tested against its self (ref-ref assays) were constructed using the variance components observed during the 2012 validation. Novartis intends to review, and if necessary, adjust the system and sample suitability criteria once sufficient historical data become available. The evaluation, and adjustment if necessary, will be based on the data generated within 12 months of the introduction of the (b)(4) method.

Reviewer Comments:

- *It appears that this current (b)(4) format is the result of simulations based on the previously available validation data. By reducing the number of doses from (b)(4) in the original format to (b)(4), the number of -----(b)(4)----- has to be used. A ---(b)(4)-----, compared to a (b)(4) model, will produce less precise and accurate RP estimates, especially when the data are not in the linear range. Although using ---(b)(4)--- per dose group (the original format) will give better intra-assay precision compared to -(b)(4)-- per dose, having (b)(4) assays with --(b)(4)-- per group in each assay makes statistical sense. That is, the inter-assay variability is likely to be larger than the intra-assay variability, and thus it is more efficient to double the number of replicate assays than doubling the number of (b)(4)-- per dose group.*
- *The use of shifting dilution range will increase the chance of getting the best linear range spanning (b)(4) consecutive dilutions. Linearity is important for the performance of a ---(b)(4)--- method. As long as the procedure for selecting the dilution range is pre-specified, potential bias can be avoided. With only (b)(4) dose levels, however, the linearity test will not be useful. Based on the dose response curves from validation data presented in Amendment 4 (response to Q1 of 4/16/14 IR, p30), the (b)(4)--- doses appear to be roughly in the linear range for 287-953, 961c, and OMV, but may not be in the linear range for 936-741.*
- *The system suitability criteria were determined using simulation. From the statistical point of view, simulation does provide useful information in the absence of sufficient amount of historical data. Novartis intends to evaluate the system and sample suitability criteria, using the data generated within 12 months of the introduction of the (b)(4) method. If further improvement of assay variability is deemed necessary by the review team, the system and sample suitability criteria need to be re-evaluated after assay improvement.*
- *The applicant uses a -----(b)(4)----- to determine whether a weighted average or semi-weighted average should be used to combine the 2 RP values to obtain the final reportable potency value for each test sample. With only 2 values and the large variability associated with the potency test, this ---(b)(4)----- test will have little power even at alpha level of (b)(4).*

- The narrower CI is considered by the applicant as conservative because of the use of the upper confidence limit as a potency specification criterion. However, the use of the upper confidence limit of potency is not an appropriate way of setting the potency specification (see reviewer's comments on the potency specification below). The weighted approach assumes no inter-assay variability. Although there is no RP value linked systematic bias, for each individual test sample, giving a heavier weight to the assay result with smaller intra-assay variability than the other assay result does not necessarily mean that the combined RP would be closer to the true value if there is inter-assay variability. The applicant uses this approach apparently because of the recommendation in -----(b)(4)----- . The -----(b)(4)----- takes a different view. Neither the ----(b)(4)----- the (b)(4) has regulatory authority in the US. However, since the ---(b)(4)----- test lacks power, particularly for this potency assay, this is a situation where using a weighted average to combine 2 RP values when the ---(b)(4)----- test is not significant can be potentially problematic. Given that the ----(b)(4)----- has regulatory authority in Europe, and having the same release test and specification in different countries may be highly desired by the applicant, this issue may need to be further investigated post approval.

(b)(4)

1 page determined to be not releaseable: (b)(4)

----- (b)(4) -----

-----.

---(b)(4)---

----- (b)(4) -----.

---(b)(4)---

----- (b)(4) -----

-----.

--(b)(4)--

----- (b)(4) -----
-----.

----- (b)(4) -----

-----.

Reviewer Comments:

- While data simulated by combining the original validation data of (b)(4) assays in (b)(4) independent runs for all possible combinations provide useful information on the performance characteristics of the assay under the new format and new mathematical model, the statistical information generated is limited by the information contained in the original validation dataset which is a relatively small dataset. It is not known whether data from the original validation adequately captured the assay variability and are representative of the assay performance in routine testing. A revalidation using data generated with the current assay format and analysis model after further assay improvement is recommended.

In response to the 9/18/14 IR requesting all testing data from lots released using the new (b)(4) method, data from (b)(4) lots were submitted in Amendment 16 on 10/2/14. The variability calculated from these (b)(4) released lots data are shown in Table 2 below.

Table 2: Estimated variability %CV of Relative Potency for (b)(4) method based on data of (b)(4) released commercial lots

Antigen	Geometric Mean RP	(b)(4) %CV from (b)(4) commercial lots
287-953	1.15	40
936-741	1.39	66
961c	1.43	38
OMV	1.41	36

Source: Reviewer's analysis using the dataset submitted to Amendment 16

The variability among this larger released lots dataset, which also includes an unknown amount of process variability, appears to be larger than the assay variability estimated from the current validation and the variability among the (b)(4) commercial lots shown in Table 1, especially for antigen 936-741.

- *Linearity, although it is included as one of the test validity criteria, provides assurance of parallel line model linearity, instead of dilutional linearity which is the type of linearity of primary interest in assay validation.*
- *The simulated assay data for evaluating relative accuracy are generated by shifting the reference curve at the target distance (i.e., the underlying curves for sample assays at different RP levels are all assumed to have 100% accuracy). Therefore, the inaccuracy in RP estimate of each simulated assay is the result of the use of the (b)(4) model fitted to part of the curve within the (b)(4)-- dilutions. As can be expected, when $RP = (b)(4)$, the sample curve tends to be in the same linear range as the reference curve, and therefore, has an average % bias of 0. As RP deviates from $(b)(4)$, the sample curve starts to deviate from linearity and parallelism within the (b)(4)--- dilution range, resulting in bias in RP estimate. Thus, the % bias calculated from the simulated data reflects only the bias that could result from the use of the (b)(4) model. Other parts of the entire assay procedure could introduce bias too.*

Specifications for ----(b)(4)----- Potency Assay

The specifications for all three recombinant protein antigens and OMV at release and end of shelf-life are:

- (1) UCL ----(b)(4)-----
- (2) RP (b)(4)

The applicant's justification for using the Upper 95% Confidence Limit (UCL) as the specification is based on non-inferiority of the tested lot to a full dose of clinically qualified reference standard that is assumed to have a relative potency of $b(4)$. In addition,

Failure to detect an RP (b)(4) (i.e., UCL (b)(4) can be simply due to large variability (which is likely to happen for this potency assay) and therefore is not a proper statistical method for demonstrating non-inferiority. Although the applicant, recognizing this problem, used an additional criterion of RP (b)(4) to offset high variability, it may not be adequate. The concern was conveyed to the applicant in the 4/16/14 IR letter.

The probability of passing a subpotent lot when not all four antigens have $RP = (b)(4)$ or when RP is between $(b)(4)$ --- can still be unacceptably high. Although the capability index showed that the current specifications are not liberal for this potency assay with high variability, this does not imply that the specifications are adequate to detect a subpotent lot. Evaluating specifications based on assay and process capability makes sense only when the variability is reasonably satisfactory. Because of the large variability, it is difficult to set specifications using proper statistics. The relatively low process capability index resulting from the additional criterion of $RP (b)(4)$ indicates a high risk of OOS for the applicant already, yet the criteria are still not adequate to detect a subpotent lot. The potency specifications should be re-evaluated after the potency assay is further improved and more data are collected.

The (b)(4) A assays are used for the determination of recombinant proteins (rp) and OMV antibody titers (anti-rp287-953, anti-rp936-741, anti-rp961c, and anti-OMV) in (b)(4)

----(b)(4)----- immunogenicity potency test. Reference standard, positive control, and test sera are -----

----- (b)(4) -----
The antibody titer of each test sample is quantified via interpolation against a reference standard curve. The quantitative procedure for determination of sample titer is different for -----(b)(4)-----.

--(b)(4)

Description of ---(b)(4)--

Novartis relied on the recommendations of the draft documents -----(b)(4)-----
----- received from CBER to develop -----

----- (b)(4) -----
-----.

System suitability criteria:

- Slope and intercept of the reference line must be within acceptable limits (2-sided 99% prediction limits of the qualified reference)
- Calculated titer for the control serum must be within acceptable limits (2-sided 99% prediction limits of the control lot used).

The (b)(4) test must be repeated for --(b)(4)--- that do not comply with at least one of the above acceptability criteria. No sample suitability criteria were provided. Upon information request on how parallelism between the reference curve and the sample curve was assessed, which is an underlying assumption for the reference line method, Novartis provided an analysis using data from the (b)(4) validation. A series of standard vs. standard and sample vs standard comparisons were analyzed to assess parallelism by (1) p-value from testing H_0 : two slopes are equal; and (2) ratio between the slopes. Of the (b)(4) comparisons for each antigen, only one comparison (sample vs standard for 961c) showed a highly significant difference between slopes (p (b)(4)). However, Novartis showed that by selecting a more appropriate dilution range in the analysis (different from the range selected by the operator), parallelism could be improved. Novartis also reported that the ratio between slopes for standard vs. standard comparison ranges from --- (b)(4) --- and from --- (b)(4) ----- for sample vs. standard comparison, well within the --(b)(4)--- range.

A serum sample is classified as “non-responder” when the titration curve has less than four consecutive points in the acceptable range for OD values, if tested at the lowest dilution exploitable (b)(4)---. The minimum values assigned to non-responder sera are:

Antigen 287-953	--(b)(4)--
Antigen 936-741	--(b)(4)--
Antigen 961c	--(b)(4)--

These minimum values were obtained by dividing the LLOQ values by 2.

Reviewer Comments:

- The details of the ---(b)(4)--- Assay” calculation method used for ---(b)(4)----- is not completely clear to me. I assume that it follows CBER’s draft documents ----- (b)(4)----- . However, a method recommended more than 20 years ago may not be up to today’s standard.
- The current SOP requires only at least one of the system suitability criteria to be met. In order to have adequate control of the assay procedure, usually all system suitability criteria need to be satisfied. There are also no criteria on the goodness of fit of the reference curve.
- Sample validity criteria need to be established to ensure the linearity and parallelism of the test sample curve. It appears that the current SOP relies on the operator to manually select the portions of the curves that are in the linear range and are parallel between the reference and test sample curves. This procedure may not be reliable enough.

Validation of ---(b)(4)---

----- (b)(4) -----

----- (b)(4) -----

----- (b)(4) -----

----- .

----- (b)(4) -----
----- .

---- (b)(4) --

----- (b)(4) -----

1 page determined to be not releaseable: (b)(4)

- According to the validation report (Tables 3.2.P.5.3.2-3, -4, -5, and -6), precision (repeatability and reproducibility) was evaluated on the OD values, not the reportable values of antibody concentration --(b)(4)--. These could be errors. Otherwise, assay parameters should be evaluated on the reportable values of the assay. Furthermore, precision was evaluated using the reference standard at (b)(4) dilutions and the positive control at a (b)(4) dilution only. The estimated variability may not represent the variability for individual test sera across a wide titer range.
- Linearity assessed is not dilutional linearity.
- LLOQs were not validated by demonstrating satisfactory precision and accuracy. No assay range was defined.

In Amendment 4, Novartis explained that a value of ½ LLOQ was assigned to non-responder sera and explained how LLOQs were determined. The following cubic model was fitted to the data from valid standard curves:

$$y = a + b_1x + b_2x^2 + b_3x^3$$

where y is the pre-assigned titer of the standard for each dilution, and x is the OD for the same dilutions calculated as the average of the values from the --- (b)(4)-- in the precision experiment.

Considering a minimum detectable OD value of (b)(4) and using the estimated model, the titer (predicted y) corresponding to (b)(4) OD (given x) can be obtained and further multiplied by the lowest dilution factors involved to obtain the LLOQ value.

Reviewer Comments:

- The LLOQs calculated by the applicant are the titers that correspond to the predetermined minimum detectable OD value of (b)(4). There was no explanation of how this minimum detectable OD value was determined. Furthermore, there was no evaluation of the assay performance at the claimed LLOQs.

OMV (b)(4)

Description of OMV (b)(4)

The OMV antibody titer in the ---(b)(4)-----is quantified via interpolation against a reference standard curve. The -----
----- (b)(4)----- of test samples is interpolated from the standard curve. All OD values of the test sample sera that are outside the estimated minimum and maximum ODs of the standard serum are eliminated. The ODs from the remaining dilutions of test sample sera are used to calculate the intermediate titers interpolated from the estimated (b)(4) standard curve. The mean and standard deviation of all available intermediate titers are calculated, and those intermediate titers outside the [mean (b)(4)] interval are excluded. If at least (b)(4) intermediate titers are still present, the mean and standard deviation are calculated for the remaining titers, and again any titers outside the [mean (b)(4)] interval are excluded. If at

least (b)(4) intermediate titers are still present, the mean of the remaining intermediate titers is then calculated. Thus, the final titer for a test sample is calculated from the combination of dilutions giving the lowest standard deviation.

System suitability criteria for OMV (b)(4):

- The average OD of the (b)(4) of the standard curve, without subtraction of the blank, is within acceptable limits for the qualified reference standard used;
- ---(b)(4)---;
- The slope of the standard curve is within acceptable limits for the qualified standard used;
- At least (b)(4) of the (b)(4) positive control sera is within the acceptable limits for the serum lot used.

The (b)(4) test must be repeated for ---(b)(4)----- that do not meet at least (b)(4) of the validity criteria listed above. In addition, at least (b)(4) of the (b)(4) controls must be within the limits of acceptability defined. No sample suitability criteria were provided. A value of ---(b)(4)--- is assigned to non-responders with ODs below the acceptable range when tested at the minimum permitted dilution.

Reviewer Comments:

- *Requiring only at least one of the validity criteria be met and at least one of the two controls be within the limits may not provide adequate control of the assay.*
- *There were no validity criteria for test sample curves. Instead, the applicant used a data elimination procedure to select the data points that gave the lowest standard deviation. Theoretically, titers obtained from dilutions of the test sample with OD values near the upper or lower asymptotes of the standard curve may be more variable and more likely to be excluded from the calculation of the final sample reportable titer value. If the sample curve is parallel to the standard curve, the standard deviation among the interpolated titers should be small. However, because there are only relatively few data points for each sample curve, it is not clear whether such a data truncation process can produce bias. It is a better practice to establish adequate system and sample suitability criteria to ensure the validity of the test.*

Validation of OMV (b)(4)

 -----(b)(4)-----

 -----.

 -----(b)(4)-----

1 page determined to be not releaseable: (b)(4)

---(b)(4)---

(b)(4)

(b)(4)

(b)(4)

---(b)(4)---

(b)(4)

---(b)(4)---

(b)(4)

- *It is not clear how the highest allowable dilutions of the test sera were determined and how they were converted to define the assay range in UE/mL.*

4.3 ENDOTOXIN SPECIFICATION

The proposed endotoxin specification for final drug product is ---(b)(4)-----.

Justification for Endotoxin Specification

(b)(4) clinical lots were tested using the revised (b)(4) assay, with $b(4)$ results obtained for each lot ($b(4)$ determinations by each of three operators). The specification was set by determining an approximation of the upper bound of a one-sided 99% prediction interval computed on the natural log (ln) scale, using the following equation:

$$\text{specification limit} = \exp (\bar{x} \pm (t_{0.01,10} \cdot \sigma_{\ln}))$$

where \bar{x} is the mean of all the data from the $b(4)$ clinical lots on the natural log scale, $t_{0.01,10}$ corresponds to the 99% probability in a one-tailed t-distribution with 10 degrees of freedom (by Satterthwaite's approximation), and σ_{\ln} is an estimate of the standard deviation of all the data on the ln (natural log) scale from the ($b(4)$) clinical lots computed as:

$$\sigma_{\ln} = \text{SQRT} (\sigma_{\text{LOT}}^2 + \sigma_{\text{Operator(Lot)}}^2 + \sigma_{\text{RESIDUAL}}^2).$$

The upper specification limit obtained from the calculation was (b)(4) IU/mL, which was rounded to (b)(4) IU/mL.

On August 22, 2014, the product reviewer sent an IR, requesting the applicant to lower this specification since the $b(4)$ lots submitted in support of the license application have endotoxin concentration that is over 2-fold lower than the specification. Novartis responded with a process capability analysis on the untransformed endotoxin data from ($b(4)$) production batches (submitted to BLA 125546.0.9 on 9/3/14). The estimated process performance index (Ppk) is ($b(4)$), with expected OOS rate of ($b(4)$), which is in line with the expected OOS rate when the 99% prediction interval was used to set the specification. Based on the low Ppk, Novartis did not consider lowering the specification to be justified.

On 9/15/2014, the product reviewer and statistical reviewer sent another IR to the applicant, requesting details on the data for $b(4)$ clinical lots and the analyses performed to justify the endotoxin specification. Justification for calculating the upper bound of the 99% prediction interval on the log scale, while the data of ($b(4)$) production batches showed no significant deviation from normality and the process capability analysis was performed on the untransformed scale, was also requested.

Novartis's responses to the 9/15/14 IR were submitted to Amendment 15 on 9/29/14. The dataset and the experimental design were provided. Each reportable value is the geometric mean of $b(4)$ replicate results. The $b(4)$ reportable values for each lot were obtained by 3 operators ($b(4)$ per operator) using $b(4)$ different ($b(4)$) lots. The estimated variance components (using the method of restricted maximum likelihood) are:

<u>Source of variability</u>	<u>Variance estimate (on ln scale)</u>
Lot	0.01399
Operator(Lot)	0
Residual	0.01102

Novartis's rationale for calculating the 99% prediction interval on the log scale is that the analysis model for the standard curve used in the (b)(4) assay is a linear model of log(reaction time) versus log(concentration). Thus, in theory, the endotoxin concentration measurement is expected to follow a lognormal distribution.

Reviewer Comments:

- *It is difficult to confirm the distribution of a measurement with limited data. Therefore, although the upper limit of the prediction interval obtained from the log transformed analysis tends to be higher than the untransformed analysis, the applicant's rationale regarding the theoretical basis for analysis on the log scale is acceptable.*
- *The estimated variance components suggest that the variability due to the manufacturing process (lot-to-lot variability) is as large as, or slightly larger than, the assay variability. The assay variability (%GCV) estimated from this dataset used to set the specification is $b(4)$, which is generally in line with the assay variability obtained from the validation (intermediate precision %GCV ranges from --- $(b)(4)$ - for different lots and different dilutions). The total variability (%CV), including the assay variability and process variability, among the $(b)(4)$ commercial lots is $b(4)$. However, the variability displayed among the stability data available in the submission appears to be quite huge. Only the data for the later stability lots appear to be less variable.*

An IR requesting additional information was sent on 12/3/2014. The main concerns were the assay variability and robustness in routine release testing and stability testing. The product reviewers also requested to have both upper and lower specifications.

In addition to the responses to the IR questions, Novartis submitted to Amendment 35 three technical reports on many experiments conducted during the assay development and optimization in order to address the questions raised in the 12/3/2014 IR. In 2012, upon observing large variability in stability data, Novartis performed an extensive evaluation and optimization of the assay, aiming to reduce the variability. As a result, the assay was modified to its current format. The validation report submitted to the BLA is for the modified method; while the stability data submitted include the data tested by the previous method, since there are limited stability data available which are initiated after the implementation of the new method in 2012. Novartis did not consider a lower limit of endotoxin level necessary.

Among the changes made to the previous assay, changing the -----(b)(4)----- has the greatest effect on reducing the assay variability. The stability data initiated since October 2012 did show a much better and consistent variability profile. The estimated variability among the available 18 months stability data tested using the modified method, assuming the residuals which could not be explained by the stability trend are due to assay variability, is 16%. This variability is comparable to that estimated in the validation of the modified method.

Reviewer Comments:

- *While there are some statistical weaknesses in the experimental designs and analyses of the numerous experiments aimed to explain the variability of the (b)(4) assay (for example, using a significance test to test an equivalence hypothesis, and testing interactions between factors with very small sample size at alpha level of 0.05, etc.) and some inconsistent results across experiments with different objectives and designs (possibly due to limitations in design and analysis), it is apparent that substantial differences in both the mean and variability between endotoxin results generated at ----(b)(4)----- are consistently observed across experiments. The ---(b)(4)----- time produces much higher endotoxin values (about 2-fold higher) with much smaller variability. In order to reduce the variability, the current assay uses a ----(b)(4)---- time. The endotoxin results should then be interpreted with this fact in mind: the ----(b)(4)----- time generates much higher values as compared to the previous method.*
- *The proposed endotoxin specification is calculated based on the data for the clinical lots which have been demonstrated to be safe in clinical trials. With the current assay and process capability, the proposed specification is not overly liberal. For quality attributes related to safety, however, the assay and process capability should not be the only factors considered in determining the specifications. Note that there is no safety signal detected in the clinical trials. However, if the product reviewers feel that a lower upper limit of endotoxin is desired, then the review team may consider whether and when additional work may be needed to either reduce the mean endotoxin level in the manufactured lots or to further improve the variability due to assay or manufacturing process.*
- *The statistical reviewer defers to the product reviewers on whether the proposed one-sided upper limit of --- (b)(4) -- is acceptable.*

5. CONCLUSIONS AND RECOMMENDATIONS

(b)(4)--- Immunogenicity Potency Test

The original potency assay, which has been used since before EU approval in 2013, has very large variability. The current immunogenicity potency test differs from the original test procedure in the assay format (using only b(4) dilutions instead of ----(b)(4)-- per group instead of --- (b)(4) -- independent assays to obtain the final reportable value instead of 1 assay) and the analysis model (parallel line analysis with shifting range (b)(4)--- instead -----(b)(4)-----). The system suitability criteria were determined by

simulation, and the validation was conducted by using the 2012 validation data for the original potency test to generate all possible potency results under the (b)(4) model. In addition, dilutional linearity and accuracy were not assessed. The variability, although improved, appears to be still quite large, based on either the simulated validation or the release data available. Because of the large variability, setting the potency specification is a challenge.

The (b)(4) assays, which are used to measure the antibody titers of the sera from ---(b)(4)----- in the potency test, generally do not have an adequate control strategy in place. The system suitability criteria appear to be less than optimal and the sample validity criteria are lacking. The validation parameters of precision, linearity, and LLOQ do not appear to have the same meanings in comparison to how they are usually defined. Therefore, it is not clear whether the (b)(4) assays could have contributed to the large variability of the immunogenicity potency test.

With high variability, the current immunogenicity potency test may only be able to detect substantial differences in potency, and therefore is recommended to be used only as an interim test until further improvements can be implemented. The current specifications are likely to be adequate to ensure that no substantially subpotent lots are released. Further revision of potency specifications will be needed after the potency assay is improved. Novartis in response to 11/12/14 IR has committed to improve the immunogenicity test and the (b)(4) assays and to revise the potency specifications (Amendment 30, dated 12/1/14).

Endotoxin Specification

The concern of the current endotoxin upper limit ---(b)(4)----- being too high was initially complicated with the uncertainty about the assay variability due to the huge variability observed in the stability data. The assay variability issue was satisfactorily resolved, from the statistical perspective, after Novartis explained that only the stability lots initiated after October 2012 are tested by the modified method, which is consistent with the more consistent and precise stability data observed in the later stability lots. The variability of the current modified (b)(4) assay is found to be consistently (b)(4) in the datasets from the validation, the clinical lots, and available commercial lots and stability data tested after the implementation of the new improved assay. The modified assay, although having much smaller variability, generates much higher values compared to the previous assay. The data for the clinical lots also showed that the variability contributed by the manufacturing process is approximately the same magnitude as the assay variability.

Based on the clinical lots data used to set the specification and the (b)(4) production lots, this upper limit does reflect the current assay and process capability. The statistical reviewer defers to the product reviewers on the acceptability of this one-sided endotoxin upper limit.