DEPARTMENT OF HEALTH AND HUMAN SERVICES
FOOD AND DRUG ADMINISTRATION
CENTER FOR DRUG EVALUATION AND RESEARCH
OFFICE OF TRANSLATIONAL SCIENCE
OFFICE OF BIOSTATISTICS

Statistical Review and Evaluation

CLINICAL STUDIES

NDA:	20-845/S_011
Name of drug:	INOmax (nitric oxide) inhalation
Indication:	prevention of chronic lung disease in preterm infants
Applicant:	INO Therapeutics LLC
Dates:	Letter June 25, 2010; PDUFA December 25, 2010
Review priority:	P
Biometrics division:	Division of Biometrics II
Statistical reviewer:	Joan Buenconsejo, Ph.D.
Concurring reviewers:	Thomas Permutt, Ph.D.
Medical division:	Division of Pulmonary, Allergy and Rheumatology Products
Clinical team:	Kimberly Witzmann, M.D.
	Anthony Durmowicz, M.D.
Project manager:	Angela Ramsey

Keywords: NDA review, clinical studies, interim analysis, multiplicity

TABLE OF CONTENTS

List of Tables

List of Figures

1 EXECUTIVE SUMMARY

1.1 CONCLUSIONS AND RECOMMENDATIONS

INOmax (nitric oxide) for inhalation received approval for hypoxemic respiratory failure due to pulmonary hypertension in the term and near-term neonate in December 1999 from the US Food and Drug Administration and subsequently in the European Medicines Agency, as well as national health authorities in Canada, Australia, Japan, and other countries.

On 30 April 2010, INO Therapeutics received a Written Request letter from the Division of Pulmonary, Allergy and Rheumatology Products, consistent with the 14 April 2008 Proposed Pediatric Study Request submitted by INO Therapeutics. The primary efficacy and safety studies outlined in the Written Request are:

- Study 1: INOT27 (EUNOS), Low Dose Inhaled Nitric Oxide (iNO) for Prevention and Treatment of Chronic Lung Disease in the Preterm Infant
- Study 2: BALLRl, Low dose inhaled nitric oxide (iNO) for prevention and treatment of chronic lung disease (CLD) in the preterm infant.

The INOT27 study is a European-based trial sponsored and executed by INO Therapeutics. The BALLRI study is a US-based National Heart, Lung, and Blood Institute (NHBLI) multicenter trial sponsored and executed by Dr. Roberta A. Ballard, conducted under IND 58,146.

On July 25, 2010, the Applicant submitted a supplemental NDA (Serial No. 11) that includes the results from the INOT27 and BALLRI studies, as part of the response outlined in the Written Request. They proposed to revise the current approved label with the following:

Current Pediatric Use Section of the INOmax Label:

Nitric Oxide for inhalation has been studied in a neonatal population (up to 14 days of age). No information about its effectiveness in other age populations is available.

Proposed Pediatric Use Section of the INOmax Label:

(b) (4)

(b) (4)

The FDA Pediatric Exclusivity Board reviewed and determined that the Applicant has 'fairly responded' to the Written Request on November 2, 2010 and therefore qualifies for Pediatric Exclusivity.

This statistical review evaluates the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study. From a statistical perspective, because of the issues surrounding the design and analyses of the BALLR1 study (discussed below), there is insufficient evidence to support the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study. (b) (4)

.

1.2 BRIEF OVERVIEW OF CLINICAL STUDIES

On July 25, 2010, the Applicant submitted a supplemental NDA (Serial No. 11) that includes the results from the INOT27 and BALLR1 studies, as part of the response outlined in the Written Request. The INOT27 study is a European-based trial sponsored and executed by INO Therapeutics. The BALLR1 study is a US-based NHLBI multicenter trial sponsored and executed by Dr. Roberta A. Ballard. INO Therapeutics LLC received permission from the Steering Committee and NHLBl to use all data from the BALLR1 and related documents for the preparation of this report and application In the Applicant's study reports, they stated that efficacy was not established in the INOT27 study. They added that the BALLRI study, "selecting for a subset of premature infants that was both viable and at risk for bronchopulmonary dysplasia (BPD), and using a higher dose exposure" was able to demonstrate significant improvement in survival without BPD.

This statistical review focuses on the BALLR1 (Ballard) study. A brief summary of the results from the INOT 27 is also presented in the review.

The BALLR1 study evaluated 587 preterm infants with a birth weight of 500 to 1250 g and between the ages of 7 and 21 days. Nitric oxide for inhalation at an initial dose of 20 ppm or placebo (l00% Grade 5 nitrogen gas) was administered using the INOvent delivery system (as defined in the protocol) as an adjunct to oxygen therapy/mechanical ventilation, and treatment was initiated within 7 to 21 days after birth. The primary efficacy variable for this study was survival without BPD at 36 weeks' PMA.

1.3 STATISTICAL ISSUES AND FINDINGS

After careful review of the application, three statistical issues were identified that warranted further consideration when interpreting the results. A brief summary of these three issues is described below. This is followed by the presentation of results from the analyses of the primary endpoint, and

lastly, a brief discussion on how these issues may have affected the interpretation of the results and conclusion. A more detailed description of the issues can be found under Section 3 of the review.

The following is a summary of the issues:

Issue #1: Data Quality

The Applicant described their SAS dataset as 'cleaned'. It is not clear what the Applicant mean by 'clean'. Conversion and transformation of legacy data to Study Data Tabulation Model (SDTM) format is often challenging particularly when the case report form was not created based on the SDTM standards. Furthermore, there are variables in the SDTM domains that are not clear and could easily be mistakenly populated. It is also possible that same variable names in two different domains will be populated differently. There may also be data collected in the CRF that do not fit in any of the SDTM domains (e.g. supplemental qualifier or not). Therefore, missing or incorrect data transformation is likely to happen. In fact, the Applicant identified missing data during transformation.

In addition, the data conversion process was conducted after publication of the results. This raises concern that the data is already unblinded and therefore may influence the outcome of the study. As an example, there is discrepancy in the results of the analyses of secondary endpoints between Ballard's NEJM report and the Applicant's report. They claimed that they used a different analysis population. I also found discrepancies in the end dates of several events (e.g. end date of mechanical ventilation) across different datasets. This again raises the concern about the quality of the data provided by the Applicant.

Issue #2: Application of different statistical tests to the primary efficacy endpoint

As stated earlier, the BALLR1 study was an NHLBI-sponsored and was executed by Dr. Roberta A. Ballard. It was only after the publication of the results did the Applicant, INO Therapeutics LLC received permission to use all the data and related documents from this study. Although Dr. Ballard submitted a protocol, she did not write a formal statistical analysis plan before the un-blinding of the data, INO Therapeutics LLC wrote their own statistical analysis plan and was finalized and signed on 14 April 2008.

In Dr. Ballard's report, different statistical tests were applied to the primary outcome variable, and the selection of some of the analysis strategies (e.g. GEE using 'exchangeable' structure and MO technique) were done after unblinding of the data and not based on formal Blind Review. The original analysis proposed by Dr. Ballard was the Fisher's exact test using all enrolled subjects. However, randomization was conducted only on first eligible infant within a family. This implies that the co-gestational siblings were not randomized and instead received the same treatment assignment as that of their sibling (first eligible infant randomized). It was only when the DSMB statistician suggested, that Dr. Ballard added the GEE model and MO technique to their collection of statistical tests to be used in the primary endpoint to account for correlation across siblings. Although applying different statistical tests potentially raises a multiplicity concern, the GEE model

and the MO technique[1] are comparable approaches. In addition, given that the primary analytical approach is wrong for this type of data, multiplicity becomes less of a problem in the original Ballard study report,

The problem of multiplicity occurs when the Applicant, knowing the issue about the cluster randomization, applied Fisher's exact test to the primary endpoint using only randomized subjects (or first eligible infant). They excluded all co-gestational siblings in their primary analysis. They conducted two additional analyses using GEE model and MO technique and were considered secondary. In this scenario, the Applicant uses three different statistical tests to analyze the primary endpoint using two different populations (i.e. all randomized versus all enrolled). Although GEE and MO techniques are comparable, these are not the same as Fisher's exact using only all randomized subjects. In this situation of multiple testing, the level of significance needs to be adjusted.

Issue # 3: Adjustments to the level of significance: Interim Analyses

According to the Applicant, two interim analyses at approximately 20% and 50% of the outcome data available were conducted. The results from the interim analyses were not reported. They claimed that the interim analysis was based on an O'Brien-Fleming stopping rule. The stopping rule was for either early rejection of the null hypothesis, if iNO is extremely efficacious or early acceptance of the null hypothesis, i.e. lack of effcacy of iNO. The exact boundaries and nominal p-values for stopping were based on applying the Lan-DeMets use function to the O'Brien-Fleming stopping rule. *They added that because these boundaries are both for rejection and acceptance of the null hypothesis, the final look at the data, considering the trial was not stopped at the interim analysis, will use a p = .05.*

The Applicant submitted their response to the Division's information request regarding the interim analyses question on October 12, 2010. They stated that they agree that the final analysis should use an alpha of less than 0.05 due to the interim analyses specified in the protocol. They provided information regarding the actual interim analyses performed at 24% and 59% of the total enrollment instead of the planned 20% and 50%, respectively. Using these proportions, they calculated the alpha levels of 0.000 for the first interim look, 0.007 for the second interim look and **0.043** for the final analysis. They also added that the results of the interim analyses were not provided by the NHLBI DSMB and only a two-page report by the lead statistician Dr. Cnaan was available.

A summary of the results from the analyses of the primary endpoint is presented in Table 1. Applying the GEE method and MO approach, there is evidence of a treatment difference between inhaled nitric oxide (iNO) and placebo on survival without BPD, even after adjusting the level of significance to 0.043. Applying the GEE method, the number needed to treat and the 95% confidence interval were also calculated. The treatment difference is about 8.9% in favor of iNO (95%CI: 0.8%, 17.1%). This implies that for every 11 patients treated with iNO, one patient will survive without BPD, with a 95% confidence interval that ranges from 1:6 to 1:125. In addition, for

[1] Follman D., Proschan M., Leifer E., "Multiple Outputation: Inference of Complex Clustered Data by Averaging Analyses from Independent Data", Biometrics (53) 420-429, 2003.

infants 7 – 14 days at enrollment, iNO showed better results than placebo in reducing the risk of being alive without CLD at 36 weeks GA. In contrast, there was no difference between treatments in the group of infants between the ages of 15 and 21 days. In the mother's race stratification, there is no numerical difference in treatment effect on "white", while treatment difference favoring iNO is seen in non-whites.

Although there is evidence of a treatment effect, data quality and possible multiplicity remains potential issues. Given that the results are known, the choice of multiplicity adjustments to use is limited. Hierarchical order or sequential testing is not suitable given the results are known, while it is unclear which statistical test methods to include in the Hochberg's procedure. On the other hand, Bonferroni approach may be too conservative. Therefore, the evidence from this study is insufficient to support the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study.

Data quality is definitely something the Applicant tried to provide us with so we can make informed assessments and decisions about their application. I also understand the efforts they made in cleaning the data and the prospective planning given they did not conduct the trial. However, I can not ignore the discrepant dates of event. The primary outcome variable (i.e. survival without BPD) relies on the actual date of 36 weeks post-menopausal age. This implies that the correct date should be used to assign the subject as a success. Quoting Dr. Woodcock's definition of data quality, "data should be good enough for a decision not to change if completely accurate data were used." In this instance, using different datasets yielded slightly different results. Therefore, like the multiplicity problem, the evidence from this study is insufficient to support the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study.

Table 1: Primary Outcome Summary – Survival without BPD

	INO	Placebo	OR	p-value
Ballard's 1^0 Analysis*	N=294	N=288		
Without	129 (44%)	105 (36%)	1.36 (0.98, 1.90)	0.0759
Sponsor's ITT†	N=269	N=268		
Without	121 (45%)	95 (35%)	1.489	0.028
GEE	N=294	N=288		
Exchangeable (reported)			1.45 (1.03, 2.04)	0.0332
Multiple Outputation Seed 1278632102	N=269	N=269		
	120 (45%)	95 (36%)	1.473	0.0370

* Fisher's Exact Test using 582 subjects
** Adjustment to stratified variables: weight strata and site
† Sponsor's ITT: includes only randomized or first siblings

2 INTRODUCTION

2.1 OVERVIEW

INOmax (nitric oxide) for inhalation received approval for hypoxemic respiratory failure due to pulmonary hypertension in the term and near-term neonate in December 1999 from the US Food and Drug Administration (FDA) and subsequently in the European Medicines Agency (EMA), as well as national health authorities in Canada, Australia, Japan, and other countries.

> The approved product, INOmax, is a vasodilator, which, in conjunction with ventilatory support and other appropriate agents, is indicated for the treatment of term and near-term (>34 weeks gestation) neonates with hypoxic respiratory failure associated with clinical or echocardiographic evidence of pulmonary hypertension, where it improves oxygenation and reduces the need for extracorporeal membrane oxygenation. The recommended dose is 20 ppm, maintained for up to 14 days or until the underlying oxygen desaturation has resolved. It must be delivered via a system which does not cause generation of excessive inhaled nitrogen dioxide.

On 30 April 2010, INO Therapeutics received a Written Request (WR) letter from the Division of Pulmonary, Allergy and Rheumatology Products, consistent with the 14 April 2008 Proposed Pediatric Study Request (PPSR) submitted by INO Therapeutics. The primary efficacy and safety studies outlined in the Written Request are:

- Study 1: INOT27 (EUNOS), Low Dose Inhaled Nitric Oxide (iNO) for Prevention and Treatment of Chronic Lung Disease in the Preterm Infant
- Study 2: BALLRl, Low dose inhaled nitric oxide (iNO) for prevention and treatment of chronic lung disease (CLD) in the preterm infant.

The INOT27 study is a European-based trial sponsored and executed by INO Therapeutics. The BALLRI study is a US-based NHLBI multicenter trial sponsored and executed by Dr. Roberta A. Ballard, conducted under IND 58,146.

On July 25, 2010, the Applicant submitted a supplemental NDA (Serial No. 11) to revise the current approved label with the following:

> **Current Pediatric Use Section of the INOmax Label:**
>
> Nitric Oxide for inhalation has been studied in a neonatal population (up to 14 days of age). No information about its effectiveness in other age populations is available.
>
> **Proposed Pediatric Use Section of the INOmax Label:**

(b) (4)

In this submission, the Applicant included the results from the INOT27 and BALLRI studies, as part of the response to the WR. They also included the results from Study INOT25, which was very similar in design to INOT27 but was not part of the response to WR. In the Applicant's reports, they stated that efficacy was not established in the INOT27 study. They added that the BALLRI study, selecting for a subset of premature infants that was both viable and at risk for BPD, and using a higher dose exposure, was able to demonstrate significant improvement in survival without BPD.

The development plan for INOmax (nitric oxide) for the treatment of chronic lung disease for premature infant was previously discussed during several meetings and correspondences with the US Food and Drug Administration's (FDA's) Division of Cardiovascular and Renal Products (DCRP) from 2004 through 2009 and with the Division of Pulmonary, Allergy and Rheumatology Products (DPARP) from 2009 onwards under IND 44,124, as well as under NDA 20-845. In previous discussions, both Divisions had questioned the validity of the primary endpoint result from the Ballard study and that a determination as to how appropriate they were could not be made without reviewing the full study report.

The main focus of this statistical review is on the BALLR1 (Ballard) study and its history. For the purpose of completion, the results from INOT27 will also be included in the review.

2.2 DATA SOURCES

All data was supplied by the applicant on CD in SAS version 9 format based on the Study Data Tabulation Model standards. The datasets were copied to EDR and available at **\\FDSWA150\NONECTD\N20845\S_011\2010-08-19.** The final study reports were submitted in paper format and archived in the document room. The information needed for this review was contained in modules 1, 2.5, 2.7, and 5.3.5 of this submission.

3 STATISTICAL EVALUATION

3.1 EVALUATION OF EFFICACY

3.1.1 BACKGROUND (BALLR1 STUDY)

The BALLR1 (Ballard) study was an investigator-initiated and conducted as a cooperative effort by the National Heart, Lung and Blood Institute (NHLBI) extramural network.

> The study was designed and initiated via a research grant from NHLBI, study number DO 1-HL625 14 and Principal Investigator IND #58,146. The Principal Investigator and the study Steering Committee designed the protocol, data collection forms, the study manual of procedures, and conducted site monitoring, serious adverse event (SAE) reporting, and other regulatory activities during the conduct of the trial. Randomization of study drug was performed centrally for all participating institutions by the DCC at the Division of Biostatistics at The Children's Hospital of Philadelphia. INO Therapeutics LLC provided study supplies and blinded INOventQ delivery devices to each institution.

The first subject was enrolled on April 4, 2000 and the last subject completed study drug at 36 weeks post-menstrual age (on July 5, 2005). During the course of the trial, the protocol was amended once, on June 2, 2004. Following completion of the study and data analysis, the results were published in The New England Journal of Medicine, July 27, 2006, Volume 355, pages 343-353; a correction was published on October 4, 2007, Volume 357, page 14. Following publication, INO Therapeutics LLC received permission from the Steering Committee and NHLBl to use all data and related documents for the preparation of this report.

Due to the non-existence of a formal statistical analysis plan prior to the un-blinding of the data, INO Therapeutics LLC wrote their own statistical analysis plan and was finalized and signed on 14 April 2008. The aim of their plan is to adhere "*as closely as possible*" to the previously defined methodology. The Applicant claimed that all statistical methods chosen to evaluate the primary and secondary efficacy endpoints are those that were specifically outlined in the June 2, 2004 protocol.

In terms of the data, the Applicant stated that

> Fully populated SAS data sets, described as "cleaned" were received from the lead investigator. The Ikaria Data Management department applied in-house designed edit checks and generated data queries. All queries were submitted to the representative of the lead investigator and/or the individual sites for resolution. All responses, as well as "Notes to File", were considered in the cleaning process and the data was appropriately modified and documented. New SAS data sets were then created from the re-cleaned data based on the Study Data Tabulation Model (SDTM) standards. These were the sets used in creating all the tables, listings and graphics for this report.

In addition, they stated that

> Following publication of the trial results by the investigators, the electronic database and electronic images of individual CRFs were retrieved by the commercial sponsor, INO Therapeutics LLC. Tabulations and analyses for this report were performed by INO Therapeutics LLC from the data

> provided by the DCC, and in accordance with the statistical analysis plan, which was created by INO Therapeutics LLC after publication and before data were evaluated. All data analysis and processing, as well as all tabulation of descriptive statistics, were performed at INO Therapeutics LLC primarily using SAS software (version 8.2 and 9.1.3) for Windows

It is not clear what "clean" means. Data transformation is often challenging particularly when the case report form was not created based on the Study Data Tabulation Model (SDTM) standards. There are variables in different domains that are not clear and could easily be mistakenly populated. There may also be variables collected that are not in any of the SDTM domains. Therefore, missing or incorrect data transformation is likely to happen. In addition, data cleaning and transformation were done after publication of the results.

3.1.2 STUDY DESIGN AND ANALYSIS

Study BALLR1 was designed as a multicenter, randomized, placebo-controlled, and blinded trial of inhaled nitric oxide (iNO) therapy (20 ppm and less) in preterm infants at high risk for developing BPD at 36 weeks' post-menstrual age (PMA). Included were subjects aged 7 to 21 days with a PMA ≤ 32 weeks and birth weight 500 to 1,250 g. Nitric oxide for inhalation or placebo (l00% Grade 5 nitrogen gas) was administered using the INOvent delivery system (as defined in the protocol) as an adjunct to oxygen therapy/mechanical ventilation, and treatment was initiated within 7 to 21 days after birth. The delivery system provides for masked delivery of the study drug. Nitric oxide was administered at an initial dose of 20 ppm for 48 to 96 hours, subsequently decreased weekly to 10 ppm, 5 ppm, 2 ppm, and finally 0 ppm. Duration of treatment was at least 24 days. Infants currently or previously enrolled in the study were eligible to receive "rescue" open-label iNO according to the criteria described under Section 9.3.4 in the study report. All infants received blinded drug and randomization were stratified by two levels of birth weight: 500 to 799 g and 800 to 1,250 g, and by site using permuted blocks of random size between 2 and 6 with equal assignments of iNO and placebo within each block. Co-gestational siblings received the same treatment assignment as the first eligible infant randomized, even if enrolled on different dates. In other words, the randomization was technically based on mothers, rather than infants, and can be referred to as 'cluster randomization.' Given that the randomization is not the usual scheme (i.e. simple randomization), conventional statistical test may not be ideal and/or applicable.

A study flow chart is presented in Figure 1.

The primary objective of the study was to examine the efficacy and safety of iNO in preterm infants of 500 to 1,250 g birth weight who continued to require mechanical ventilation or nCPAP between 7 and 21 days of age. This was more explicitly described by the first specific aim defined in the protocol:

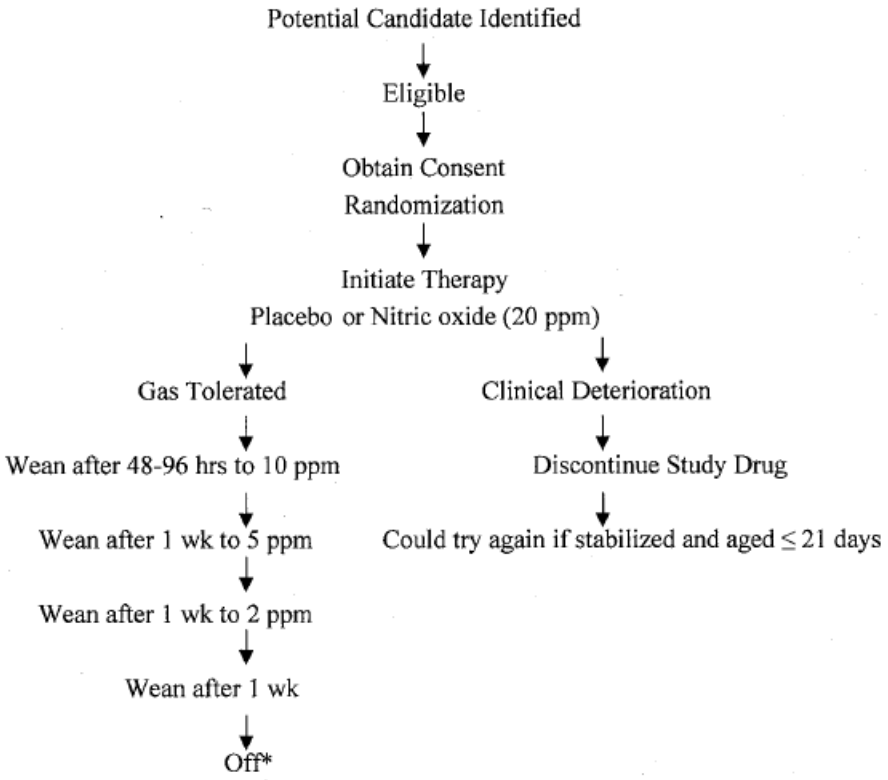
> *Specific Aim # 1: To assess the effect of iNO on the occurrence and severity of chronic lung disease in the preterm infant.*

The secondary objectives were explicitly described by the remaining specific aims described in the protocol:

Specific Aim #1: To assess airway resistance, length of ventilation and hospitalization, and decreased duration of oxygen requirement.

Specific Aim #2: To examine the safety of iNO in preterm infants at risk of chronic lung disease.

Figure 1: Randomization and Subject Data Flow



* If initially fails to come off 2 ppm, wean once a week until successful.

Source: BALLR1 Clinical Study Report, Module 5, vol. 32, page 26)

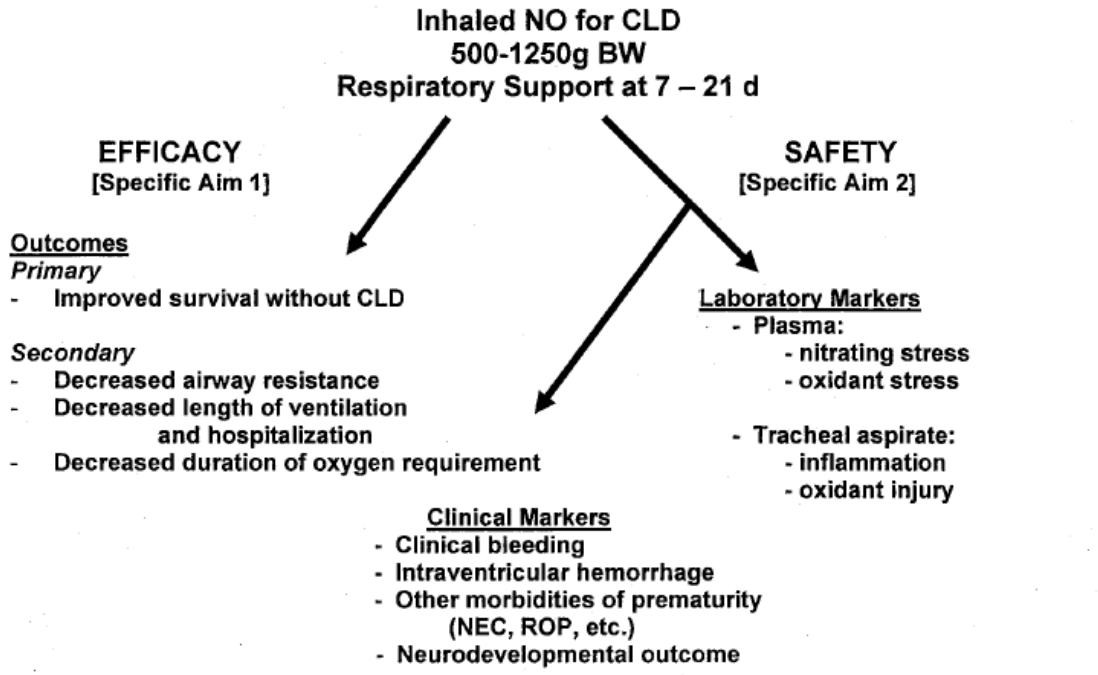
The primary efficacy variable for this study was survival without BPD at 36 weeks' PMA (Figure *2*). Bronchopulmonary dysplasia (BPD) was defined by the physiologic definition of Dr. Michele Walsh. Oxygen challenge tests were done to establish the physiologic definition of BPD. An infant who was alive without BPD at 36 weeks' PMA was counted as a "success." An infant who died, or who had BPD at 36 weeks' PMA, was counted as a "failure."

In the NEJM article, infants who required ventilatory support or were unable to maintain oxygen saturation above 88% while breathing room air were classified as having BPD.

Meanwhile, according to the NO CLD protocol, to standardize assessment of respiratory status at 36 weeks PMA, the following definitions will be used:

- All infants still requiring mechanical ventilation or CPAP are defined as having chronic lung disease (CLD).
- Infants requiring continuous effective supplemental oxygen (> 30% oxygen) via hood or nasal cannula are defined as having CLD,
- Infants requiring continuous effective supplemental oxygen ≤ 30% will undergo an oxygen reduction challenge test to determine CLD status.

Figure 2: The Outcomes of the Study (dated June 2, 2004)



Source: The NO CLD Protocol and Manual NHLBI Study # U01-HL62514, Dated June 2, 2004, page 11 and 23

The primary analysis was planned to compare the rates of survival without BPD between the iNO group and the placebo group using a Fisher's exact test. All randomized subjects were to be included in this approach.

Additional analyses were proposed in the June 2, 2004[2] protocol including

- Choosing at random one subject from each set of multiple births, and recalculating the Fisher's exact test with that group. This was suggested by the DSMB statistician (module 5, vol 32, pg 43).

- An analysis using GEE was to be performed, clustering twins and triplets, and was to include all infants enrolled. In this analysis, a logistic link was to be used, with the same outcome definition and

[2] The study protocol was amended on June 2, 2004.

treatment assignment as the explanatory variable. The type of structure used is 'unstructured' correlation matrix[3].

- An analysis using data for stratification factors was also planned. Thus, the creation of a data set of stratified 2x2 contingency tables was planned as well as use of the Breslow and Day homogeneity test to explore interactions with the stratification variables.

They also added that "regardless of whether the study is positive or not", they plan to conduct analyses using logistic regression and GEE that includes several predictors: PMA, CRIB scores, antenatal maternal corticosteroids, as well as demographic variables.

According to the Applicant, two interim analyses at approximately 20% and 50% of the outcome data available were conducted. The results from the interim analyses were not reported. They claimed that the interim analysis was based on an O'Brien-Fleming stopping rule. The stopping rule was for either early rejection of the null hypothesis, if iNO is extremely efficacious or early acceptance of the null hypothesis, i.e. lack of effcacy of iNO. The exact boundaries and nominal p-values for stopping were based on applying the Lan-DeMets use function to the O'Brien-Fleming stopping rule.

They added that because these boundaries are both for rejection and acceptance of the null hypothesis, the final look at the data, considering the trial was not stopped at the interim analysis, will use a p = .05.

It is not clear from the June 2, 2004 protocol and from the April 14, 2008 SAP what the exact boundaries and nominal p-values for stopping were. Given that the Applicant conducted two interim analyses with the purpose of stopping the trial, the final analysis should use an alpha of less than 0.05.

An information request (IR) was faxed to the Applicant on September 29, 2010 about their interim analyses. In the letter, we requested detailed information regarding the boundaries and the nominal p-values used for stopping in each interim look, as well as the level of significance for the final analyses. We also requested the results from the interim analyses.

The Applicant submitted to the Division their response to the interim analyses question on October 12, 2010. They stated that they agree that the final analysis should use an alpha of less than 0.05 due to the interim analyses specified in the protocol. They provided information regarding the actual interim analyses performed at 24% and 59% of the total enrollment instead of the planned 20% and 50%, respectively. Using these proportions, they calculated the alpha levels of 0.000 for the first interim look, 0.007 for the second interim look and **0.043** for the final analysis.

They also added that the results of the interim analyses were not provided by the NHLBI DSMB and only a two-page report by the lead statistician Dr. Cnaan was available (Appendix 1). It appears from this report that they also recalculated the sample size based on the result from the interim analysis.

[3] Ballard Study Report, 16.1 (Module 5, Vol 33, page 0417)

The Applicant also evaluated three secondary time-to-event efficacy outcomes:

- duration of ventilatory support
- duration of supplemental oxygen
- duration of hospitalization.

All three outcomes were determined by the start date of treatment and the end date/time of each variable. They were analyzed using survival analysis methods. The event in the survival analysis was to be end of ventilatory support, oxygen, and hospitalization, respectively. An observation was to be considered censored if at the end of observation (36 weeks) the event of interest (end of ventilatory support, etc.) had not occurred. For infants who died, two approaches were planned: (1) censor the observation at the time of death, and (2) the more conservative approach censor the observation at the 36 weeks' PMA point. Stratified log rank tests were to be conducted at the $p = 0.01$ significance level. The type 1 error of 0.01 was chosen and specified in the June 2, 2004 protocol to account for the multiplicity of secondary outcomes. In addition, exploratory Cox regression models were planned to examine whether the results of the time to events comparisons between the two groups are different for different PMAs, sex, race, and the stratification factors.

A fourth secondary outcome, decreased airway resistance, was also evaluated. Measurement was done in selected sites only. The differences in airway resistance before and immediately after initiation of treatment will be compared between the two treatment groups using a two-sample t-test.

Of note, all the analyses conducted for the secondary endpoints did not adjust for correlation due to family relations (i.e. siblings). I conducted additional analyses to account for correlated data.

The sample size was revised downward in March 2004 (and was part of the June 2, 2004 amendment) from the previous calculations of N = 726 and N = 690 to N=585. Three reasons are documented as the rationale for the sample size reduction:

- lower rate (7%) of multiple births than the estimated 17% rate
- improvement of statistical software allowing for sample size calculations based on spacing of two interim looks at 20% and 50% rather than 33.3% and 66.7%
- a request to decrease the power to 80% from the 85% power suggested by the DSMB prior to opening the majority of the sites (DSMB minutes March 2004)

> The calculated sample size of 585 randomized infants had an 80% power to detect this difference in proportions while controlling for a two-sided α of 0.05, allowing for two interim analyses at approximately 20% and 50% of the outcome data available.

INO Therapeutics LLC statistical analysis plan was finalized and signed on 14 April 2008. However, during preparation of the study report, the Applicant became aware of the population definition Dr. Ballard used, and the analytical approaches that were not part of the original analysis plan (e.g. multiple outputation method). They also identified missing data during transformation.

In summary, the following is a list of changes made from the original protocol:

1. The sample size was revised downward in March 2004 (and was part of the June 2, 2004 amendment) from the previous calculations of N = 726 and N = 690 to N=585. This was due to lower rate of

multiple births than what was estimated (i.e. actual rate is 7% compared to estimated 17%). The Applicant stated that this was adjusted without knowledge of the treatment outcomes and therefore was no inflation of the type 1 error.

2. The manner of treatment assignment for co-gestational infants by group instead of individuals. It was determined that the primary statistician randomized these siblings as a group, not as individuals. All patients in the group would receive the next available treatment in the randomization scheme. One sibling was designated as the "first" sibling and received the next available treatment on the list; the subsequent siblings received the same treatment as the first, but this was not necessarily the next treatment on the schedule. It was possible that, depending on the block size (which was randomly chosen), up to 3 rows in the randomization schedule could have been skipped in order to ensure that siblings received the same treatment.

3. This resulted in two ITT populations, all enrolled and randomized ITT. Therefore, only the randomized sibling was considered part of the ITT population in the evaluations contained in this report, i.e. *the ITT population is the group of individually randomized subjects, including the first enrolled of the co-gestational siblings*. However, certain evaluations were performed on the entire group of subjects including all co-gestational siblings; this group is referred to as "*all enrolled*." The protocol stipulated that all randomized subjects would be included in the primary analyses.

4. Published study results described a statistical analysis with multiple outputation (MO) technique to adjust for clustered data newly described at the time of publication and suggested by the DSMB statistician.

5. Missing data were identified during the preparation of the new SAS data sets based on the SDTM standards in preparation of this report resulting in fewer analyses and tables than described in the statistical analysis plan dated 14 April 2008.

A list of changes and omissions made by the Applicant was reported in the Study report (see Appendix 2).

3.1.3 DISPOSITION OF PATIENTS, DEMOGRAPHY AND BASELINE CHARACTERISTICS

A total of 1555 infants with a birth weight of 500 to 1250 g and between the ages of 7 and 21 days were eligible to participate in the study. Of these, 587 were randomized and enrolled between May 2000 and April 2005 (296 infants were in the iNO group and 291 infants were in the placebo group).

Three infants were withdrawn after randomization but before receiving treatment (1 iNO, 2 placebo). Of the 584 infants who received study medication, 295 were in the iNO group and 289 were in the placebo group. In Dr. Ballard's article, she reported that two withdrew after gas was started (1 iNO, 1 placebo) and she used 582 patients in all her analyses.

Patient disposition for all subjects enrolled is summarized in Table 2. Only 7% did not complete the study and they are generally well-balanced across treatment groups. As noted in Section 3.1.2, the Applicant re-defines the intent-to-treat population by applying the 'ICH E9' definition of 'randomized'. Therefore, only the randomized sibling was considered part of the ITT population in the evaluation, and does not include the cogestational siblings. Of the 587 subjects, only 542 subjects met this criterion. A summary of the subject enrollment is presented in Table 3.

Table 2: Patient Disposition – All Subjects Enrolled

	Inhaled NO N=296	Placebo N = 291	Total N = 587
Completed	278 (94%)	270 (93%)	548 (93%)
Did not complete	18 (6%)	21 (7%)	39 (7%)
Death	16	18	34
Consent Withdrawn	1	2	3
Delivery Device Failure	0	1	1
Other	1	0	1

Source: BALLR1 Clinical Study Report, Table 3

Table 3: Summary of Subject Enrollment

	Inhaled NO N=296	Placebo N = 291	Total N = 587
All Subjects who took drug and have safety data	295 (100%)	289 (99%)	584 (~100%)
All Subjects Randomized*	271 (92%)	271 (92%)	542 (92%)
have primary efficacy variable data	269	268	537
never took the drug and/or without primary efficacy data	2	3	5
Subjects who are cogestational siblings, not randomized	25 (8%)	20 (7%)	45 (8%)

Source: BALLR1 Clinical Study Report, Table 5
* Applicant's definition of ITT.

The Applicant reported several protocol violations including violation of inclusion or exclusion criteria, deviation from study guidelines, study gas dosage error, dosing schedule not followed, etc. Of the 587 enrolled, 438 subjects (75%) have at least one protocol violations, majority of which are due to deviation from study guidelines. According to the Applicant, these violations did not affect the overall conclusion.

Only four subjects were noted to have violated the inclusion/exclusion criteria (3 iNO group, 1 placebo), and four subjects were noted to have been randomized to one study drug, but initially received the opposite treatment ((3 iNO group, 1 placebo). These patients were analyzed as randomized. Subjects were eligible to receive "rescue" open-label iNO therapy who met the protocol-specified criteria. A total of 33 (6%) of 584 subjects received open-label iNO, 15 (5 %) of 295 iNO subjects and 18 (6%) of289 placebo subjects.

Demographic characteristics of subjects at baseline were generally well balanced across treatment groups for infants and mothers (Appendix 3 and Appendix 4, respectively). The treatment groups were comparable between iNO and placebo groups for PMA age, sex, age of mother, birth weight, delivery method, respiratory severe score strata, Apgar scores at 1, 5, and 10 minutes, baseline ventilation and multiple births. Fifty four percent of infants were male, 52% are Caucasian, 73% are

singletons, and mean PMA age is 28 weeks. Of note, roughly 66% of the respiratory severity scores were missing, and according to Dr. Witzmann, this further proves that the data collection is suboptimal.

The Applicant noted that 45 subjects of multiple births (8% iNO vs. 7% placebo) received study drug but were considered "non-randomized" because the method of treatment assignment did not strictly conform to the ICH E9 definition of "randomized." (Table 3) Comparing all subjects enrolled, the proportion of subjects from multiple births was similar between the treatment groups (24% iNO vs. 23% placebo, Table 4).

Table 4: Mothers with Multiple Births (All Enrolled Subjects)

	Inhaled NO N=296	Placebo N = 291	Total N = 587
Singleton	211 (72%)	214 (74%)	425 (73%)
Twins	63 (21%)	62 (21%)	125 (21%)
Triplets	10 (3%)	6 (2%)	16 (3%)
Quadruplets	0	1	1
Missing	11 (4%)	6 (2%)	17 (3%)

Source: BALLR1 Clinical Study Report, Table 9

The number of siblings included in the Safety Population is presented in Table 5. This can also be described as 'cluster'. There are 499 clusters of one, 37 clusters of two, and 4 clusters of three, for a total of 539 clusters.

Table 5: Number of Siblings included in the Trial (Safety Population)

	Inhaled NO N=295	Placebo N = 269	Total N = 584
Single	270 (92%)	269 (93%)	539 (92%)
Twins	22 (7%)	19 (7%)	41 (7%)
Triplets	3 (1%)	1	4 (1%)

3.1.4 PRIMARY EFFICACY ENDPOINT

Two interim analyses were conducted by the Investigator. The Applicant maintained that all primary efficacy analyses should be evaluated at a level of significant of 0.05. Based on IR response, it appears that the final analysis should be evaluated at the alpha level of 0.043.

The primary endpoint was analyzed in several ways by Dr. Ballard and by the Applicant. The protocol in June 4, 2004 clearly stated that the primary analysis is Fisher's exact test and will be conducted on all enrolled subjects. However, it is duly noted that this may not be the ideal analytic approach for this type of data (with clustered randomization), since this approach utilizes all subjects enrolled with efficacy data available and treat them as independent. Therefore, Dr. Ballard and the Applicant also used two additional methods to analyze the primary endpoint, multiple outputation[4] (MO) and generalized estimating equation (GEE). These two approaches required complete, intact sibling groups in order to be employed.

The MO technique requires randomly selecting one subject from each sibling group from the total population of 582 (Dr. Ballard). I re-analyzed the data using the safety population with the clusters presented in Table 5. Using uniform distribution as the random number generator with seed 1278632102, the MO technique randomly selects from the clusters to get 539 data points. This was done 1000 times and the results will be the average of the 1000 simulations. Based on literature search, this technique has not been extensively used.

The GEE is an approach used for analyzing correlated data such as those derived from related individuals. One of the important components of this approach is the choice of correlation structure between observations on the same cluster. The unstructured correlation structure was chosen by Dr. Ballard (reference: June 2, 2004 protocol) and by the Applicant's in the Statistical Analysis Plan dated April 14, 2008 (reference: Module 5, Vol. 33, page 417). However, the result presented in the NEJM and in the Study Report appears to be done using the exchangeable model.

In the IR dated September 29, 2010, we questioned the Applicant regarding the correlation matrix used. In their response, they claimed that the revised protocol dated June 2, 2004 did not specify which correlation structure to use for the GEE model, but in the Statistical Report, they did specify the designation of unstructured for the working correlation matrix. According to them, unstructured is the default mode prior to analyzing the data. They added that when they performed the GEE analysis, they also assessed the model fit by utilizing the quasi-information criterion (QIC) value to choose the correct working correlation structure. They found that exchangeable has the smallest QIC value and therefore chose to use that as their final working correlation structure.

The Applicant re-defines the intent-to-treat population by applying the ICH E9 definition of randomized. Therefore, in their primary analysis, they only included the randomized sibling as part

[4] Follman D., Proschan M., Leifer E., "Multiple Outputation: Inference of Complex Clustered Data by Averaging Analyses from Independent Data", Biometrics (53) 420-429, 2003.

of the ITT population in the evaluation, and did not include the cogestational siblings. Of the 587 subjects, only 542 subjects met this criterion. Of the 542, only 537 have primary efficacy data. A summary of the primary outcome results is presented in Table 6.

As noted, the primary approach used by Dr. Ballard (using Fisher's exact test on all enrolled subjects) and by the Applicant (using Fisher's exact test on all randomized subjects) is not ideal. Although estimates of effect from their analyses may be accurately derived from these clustered data without adjusting for correlation, but the variability of these effects would likely be biased (magnitude and direction of bias are unknown), leading to possible incorrect test statistics and confidence intervals. Therefore, appropriate statistical analysis needs to take correlation into consideration, like GEE.

Although choosing the working correlation structure was done post-hoc, it does make sense to use the 'exchangeable' structure for clustered data given that the number of siblings is small, and most likely sibling responses are correlated. In the exchangeable correlation matrix, a single correlation parameter is associated with this structure, which means that each cluster is assumed to be internally correlated in a similar manner, while subjects are assumed to be independent[5]. In contrast, unstructured correlation matrix assumes all correlations are different.

Highlighted in Table 6 is the result from the exchangeable model. Treatment difference on survival without BPD is 8.9% higher in the iNO group compared to the placebo group (applying the GEE model). The confidence interval for the odds ratio is above 1, suggesting positive association between treatment with iNO and survival without BPD.

Different statistical tests were applied to the primary outcome variable, and the selection of an analysis strategy (e.g. GEE using exchangeable structure and MO technique) was done post-hoc and not based on formal Blind Review. Although applying different statistical tests potentially raises a multiplicity concern, the GEE model and the MO technique are comparable approach[6]. In addition, given that the primary analytical approach is wrong for this type of data, multiplicity becomes less of a problem in the original Ballard study report,

The problem of multiplicity occurs when the Applicant, knowing the issue about the cluster randomization, applied Fisher's exact test to the primary endpoint using only randomized subjects (or first eligible infant). They excluded all co-gestational siblings in their primary analysis. They conducted two additional analyses using GEE model and MO technique and were considered secondary. In this scenario, the Applicant uses three different statistical tests to analyze the primary endpoint using two different populations (i.e. all randomized versus all enrolled). Although GEE and MO techniques are comparable, these are not the same as Fisher's exact using only all randomized subjects. In this situation of multiple testing, the level of significance needs to be adjusted. Given that the results are known, the choice of multiplicity adjustments techniques to use

[5] Hardin JW and Hilbe JM, "Generalized Estimating Equations" Wiley Encyclopedia in Clinical Trials 2008

[6] Follman D., Proschan M., Leifer E., "Multiple Outputation: Inference of Complex Clustered Data by Averaging Analyses from Independent Data", Biometrics (53) 420-429, 2003.

is limited. Hierarchical order or sequential testing is not suitable given the results are known, while it is unclear which statistical test methods to include in the Hochberg's procedure. On the other hand, Bonferroni approach may be too conservative. Therefore, the evidence from this study is insufficient to support the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study.

Table 6: Primary Outcome Summary – Survival without BPD

	INO	Placebo	OR	p-value
Ballard's 1^0 Analysis*	N=294	N=288		
Without	129 (44%)	105 (36%)	1.36 (0.98, 1.90)	0.0759
With adjustment**			1.36 (0.98, 1.90)	0.0679
Sponsor's ITT†	N=269	N=268		
Without	121 (45%)	95 (35%)	1.489	0.028
With adjustment				
GEE	N=294	N=288		
Unstructured (pre-specified)	129 (44%)	105 (36%)	0.27 (0.03, 2.60)	0.2553
Exchangeable (reported)			1.45 (1.03, 2.04)	0.0332
Exchangeable (calculated)			1.45 (1.03, 2.04)	0.0329**
Multiple Outputation Seed 1278632102	N=269	N=269		
	120 (45%)	95 (36%)	1.473	0.0370

* Fisher's Exact Test using 582 subjects
** Adjustment to stratified variables: weight strata and site
† Sponsor's ITT: includes only randomized or first siblings

Subjects were eligible to receive "rescue" open-label iNO therapy who met the protocol-specified criteria. A total of 33 (5.7%) of 584 subjects received open-label NO, 15 (5%) of the 295 iNO subjects and 18 (6%) of the 289 placebo subjects. Excluding subjects who received open-label iNO as rescue therapy or classifying them as nonresponders (did not survive without BPD) did not alter the conclusion using GEE model. Treatment difference on survival without BPD is 9% higher in the iNO group compared to the placebo group (Table 7).

Table 7: Primary Outcome Summary – Survival without BPD

	INO	Placebo	OR	p-value
GEE (exchangeable)				
Excluding subjects who received rescue open-label iNO	129/279 (46%)	104/270 (39%)	1.46 (1.03, 2.07)	0.0346
Subjects who received rescue open-label iNO are nonresponders	129/294 (44%)	104/288 (36%)	1.45 (1.03, 2.04)	0.0327

3.1.5 KEY SECONDARY EFFICACY ENDPOINTS

The following are the key secondary endpoints the Applicant proposed to evaluate and report.

- duration of ventilatory support
- duration of supplemental oxygen
- duration of hospitalization
- decreased airway resistance

Only the results from the analyses of the first three endpoints were reported in the Study Report.

In the IR dated September 29, 2010, we questioned the Applicant their rationale for not reporting the result from the analysis of airway decreased resistance. In their response, they claimed that measurement of airway resistance in infants is technically difficult. They added that the variable was collected only at selected sites and was not part of the Case Report Form collection. They also claimed that Ikaria considered this investigation as exploratory and did not provide the main data set to the Applicant. The Applicant added that the result of the investigation was published in 2007[7]. The investigator stated that the number of infants entered was small (71 in total), there was a significant drop out over time (26 had measurements at 2 weeks follow-up), and they concluded that there was no effect on iNO on airway resistance or lung compliance, and believed that further investigation is not warranted.

Upon discussion with Dr. Witzmann regarding the rationale of the Applicant, she agrees with the Applicant about the difficulty in measuring airway resistance in infants. Given there is a lack of information in the literature regarding this endpoint or an appropriate measurement, we deemed that this endpoint should be considered 'exploratory'.

To analyze the time-to-event endpoints, two approaches were planned for censoring: (1) censor the observation at the time of death, and (2) the more conservative approach censor the observation at the 36 weeks' PMA point. According to the Applicant, by censoring the data with the use of date of event rather than at 36 weeks' PMA, a higher percentage of events were available for analysis. They claimed that the use of the date of event allowed for a clearer picture of the estimated treatment difference due to the larger sample size and they considered it the preferred method of censoring. The type 1 error of 0.01 was chosen and specified in the June 2, 2004 protocol to account for the multiplicity of secondary outcomes.

The results from Dr. Ballard's NEJM report and Applicant's Study Report are presented in Table 8. Although the conclusions reported in the NEJM article is the same as that in the Applicant's study report, I noted that the values are different. Careful review of the Applicant's data (i.e. their efficacy data and their SDTM) suggests a multitude of discrepancies in the date of selected events (i.e.

[7] Di Fiore et al, "The effect of inhaled nitric oxide on pulmonary function in preterm infants" Journal of Perinatology, 2007, 27, 766-771

mechanical ventilation and oxygenation) that may have led to different results from the Study report (Table 8 and Table 9).

Table 8: Secondary Outcome Variables Summary – Time-to-Selected Events (in median, weeks) – Randomized Subjects

	INO N=269	Placebo N=268	p-value
End of Hospitalization			
Study Report	12.0	12.9	0.009
Ballard (NEJM)			0.04
End of Oxygenation			
Study Report	10.7	11.9	0.0043
Ballard (NEJM)			0.006

Table 9: Reviewer's Re-analyses Secondary Outcome Variables Summary – Time-to-Selected Events (in median, weeks) – Randomized Subjects

	INO N=269	Placebo N=268	HR (95% CI)*	p-value
Mechanical Ventilation				
Number of Subjects	150 (55%)	124 (46%)	1.7 (1.3, 2.1)	
Median	11.1	13.0		<0.0001
End of Hospitalization				
Number of Subjects	248 (92%)	248 (93%)	1.3 (1.1, 1.6)	
Median	12.5	13.3		0.0093
End of Oxygenation				
Number of Subjects	162 (60%)	142 (52%)	1.4 (1.1, 1.8)	
Median	10.3	11.4		0.0055

*Adjusted by gestational age, sex, and weight strata

According to the response in the IR letter, the discrepancies could either be the result from the different population selected, or from the different calculations for the duration of hospitalization, or both. Of note, the NEJM article utilized efficacy population of 587 subjects while the Clinical Study Report analyzed from 542 subjects. They also explained how they calculated the duration of ventilation, oxygenation, and hospitalization. In their approach, they considered all possible values from multiple sources (eg. DS.xpt, PS.xpt, RS.xps, SUPPRS.xpt) in the CRF, and chose the latest possible end date to populate the 'end date' in the EFF (or the efficacy analysis) dataset. Duration

was then calculated by subtracting the 'end date' with the "date of randomization". There were missing end dates in the EFF dataset either due to patient death, or missing on the original CRFs.

This is disconcerting because 'date' is an important component in defining the primary efficacy outcome. Recall that the primary outcome was survival without BPD at 36 weeks' PMA. This implies that the correct date should be used to assign the subject as 'survived or success'. It is not clear from the report or from the Applicant's explanation whether they used the same 'dates' as Ballard. This is an important issue to consider.

They also stated in the letter that they did not account or adjust for the possible correlation that may exist within sibling groups "because in the log-rank test, only one sibling in a sibling group was randomized." Therefore, the Applicant only used the randomized subjects (i.e. the first siblings) and Dr. Ballard appeared to assume that the subjects are independent.

I am in agreement that only the randomized subject should be included in the log-rank test. However, there is also an alternative approach in calculating and comparing the hazard ratios between treatment groups when data is clustered. The modified Cox proportional hazards model (SAS: PROC TPHREG) uses the sandwich estimator for the variance of the hazard ratio for correlated data.

Applying sandwich estimator for the variance of the hazard ratio to account for correlated data, the difference in duration of discharge did not meet the pre-specified level of significance of 0.01 (Table 10). In addition, given that the results from the Ballard report and from using different datasets are different, it is not clear whether there is evidence of a difference in end of mechanical ventilation or end of oxygenation between the treatment groups.

Table 10: Reviewer's Re-analyses of Secondary Outcome Variables– Time-to-Selected Events (in median, weeks) – Safety Population*

	EFF.XPT		DISP.XPT	
	HR (95% CI)	p-value	HR (95% CI)	p-value
Mechanical Ventilation	1.6 (1.3, 2.1)	0.0001	1.5 (1.2, 1.9)	0.0010
End of Hospitalization	1.2 (1.0, 1.5)	0.0168	1.2 (1.0, 1.5)	0.0168
End of Oxygenation	1.4 (1.1, 1.7)	0.0074	1.3 (1.1, 1.7)	0.0135

*Safety Population: N=582
Applicant's pre-specified $\alpha = 0.01$ to account for multiplicity

3.1.6 EFFICACY CONCLUSIONS

After careful review of the application, several issues were identified. Some where addressed by the Applicant in their response to the Information Request and some where addressed by reanalyzing the data. Three statistical issues were identified that warranted further consideration when interpreting the results. The three issues are:

1. Data Quality
2. Multiplicity – Application of different statistical tests to the primary endpoint
3. Interim Analysis – level of significance adjustment

Given that this study was originally investigator-initiated, the protocol may have been written to satisfy a research hypothesis and not intended for marketing application. The standard we used in research may be not as stringent, albeit should be, compared to regulatory standard.

In summary, the difference between treatment groups is about 8.9% in favor of iNO (95%CI: 0.8%, 17.1%). This implies that for every 11 patients treated with iNO, one patient will survive without BPD, with a 95% confidence interval that ranges from 1:6 to 1:125. In addition, for infants 7 – 14 days at enrollment, iNO showed better results than placebo in reducing the risk of being alive without CLD at 36 weeks GA. In contrast, there was no difference between treatments in the group of infants between the ages of 15 and 21 days. In the mother's race stratification, there is no numerical difference in treatment effect on "white", while treatment difference favoring iNO is seen in non-whites.

However, data quality and possible multiplicity remains potential issues. Therefore, the evidence from this study is insufficient to support the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study.

3.1.7 SUMMARY OF THE INOT27 STUDY RESULTS

Study INOT27 was a prospective, multicenter, double-blind, placebo-controlled, randomized trial of nitric oxide therapy in preterm infants with respiratory distress. The primary objective was to assess the safety and efficacy of inhaled nitric oxide (iNO) 5 ppm to reduce the risk of chronic lung disease (CLD) in preterm infants with respiratory distress compared to placebo, and to assess the long-term effects of the therapy on the development of these children over 7 years of clinical follow-up. The primary endpoint was success, defined as infant alive without bronchopulmonar dysplasia (BPD) at 36 weeks' gestational age (GA). This was analyzed on all randomized subjects using logistic regression. It is not clear from the review of the Clinical Study Report whether co-gestational siblings were included in the study and whether they were treated independently or dependently (i.e. same treatment assignment similar to Study BALLR1). The only place this was mentioned is in Section 9.8 of Study Report

> All methods outlined in the SAP dated and signed 30 June 2008 were followed with the exception of the multiple births analysis; this analysis was not performed.

In the Statistical Analysis Plan dated May 31, 2005, it outlined one interim analysis and this will occur after the first 400 the patients randomized have completed the 36 week GA assessment 36 week GA assessment.

> The trial will be stopped early if one of two conditions is satisfied. First, there is a one-sided unadjusted p-value in favor of iNOmax of 0.00153. In that case no further patients will be recruited, those already recruited will complete their treatment and an analysis of all results will be made once these further patients have completed. Second, the trial will be stopped for futility if there is a one-sided p-value of 0.10 in favor of the control treatment at the interim analysis. In all other circumstances, the trial will continue to recruit patients and will proceed to completion and an analysis of all results will be carried out at the end of the trial. In that case, the results will be deemed significant if the one-sided p-value in favor of iNOmax is less than 0.0245 (or two-sided 0.049).

A total of 800 subjects were enrolled and randomly assigned to study treatment (399 iNO and 401 placebo) and 792 subjects were included in the Safety population. Of the 792 subjects in the Safety population, 85% completed the study according to the protocol (86% iNO and 85% placebo). A total of 116 subjects (15%) did not complete the study, of which 64 were due to death (33 iNO and 31 placebo).

In general, the iNO and placebo groups were well matched for demographics and baseline characteristics in GA, birth weight, sex, Apgar scores, oxygenation index, and race. Inhaled NO-treated infants were slightly lighter and younger on average, and there were more iNO-treated infants born at less than 26 weeks GA.

In this study, treatment with iNO 5 ppm and less for 21 days was comparable with placebo ($p = 0.7340$) for survival without BPD in preterm infants with respiratory distress. There were 258 (65%) successes in the iNO group and 262 (66%) in the placebo group. For a subgroup of subjects with at least 21 days of treatment, a numerically higher success rate was observed in the iNO group compared to placebo. No treatment difference was observed for the primary outcome by BPD severity; although there is numerically higher proportion of deaths in the iNO group compared to placebo. The Applicant reported that there were no significant differences between treatment groups for the secondary variables of length of mechanical ventilation, and length of hospitalization.

The Applicant concluded that this was a failed study and added that the reasons for the failure are not completely clear. They also added that the study results are not at all dissimilar to those of Kinsella et al[8] who treated a similar study population with a similar treatment regimen; thus, "these results are (in that sense), understandable and credible." Therefore, they concluded that early use of iNO at 5ppm in preterm infants does not reduce the incidence of BPD or brain injury.

[8] Kinsella JP et al. "Early Inhaled Nitric Oxide Therapy in Premature Newborns with Respiratory Failure", NEJM 355:354-364, 2006

Table 11: Primary Outcome Summary – Survival without BPD – All Randomized Subjects

	INO N=395	Placebo N=400	OR*	p-value**
Overall: Success	258 (65%)	262 (66%)	1.05	0.7340
Subjects with at least 21 days	N=140	N=148		
Success	98 (70%)	90 (61%)		0.0839
By Severity				
Alive and no BPD because: 1) Subject Breathing Room Air, or 2) Subject Not on O2 or CPAP and did not qualify for an ORT	244 (62%)	239 (60%)		
Alive and no BPD because subject received an ORT and did not have BPD	14 (4%)	23 (6%)		
Alive and had BPD because BPD confirmed by an ORT	44 (11%)	47 (12%)		
Alive and had BPD because: 1) subject still on mechanical ventilation or 2) subject still receiving O2 or CPAP	37 (9%)	49 (12%)		
Died	56 (14%)	42 (11%)		

* iNO vs. placebo, adjustment to stratified variables: gestational age and center
** Based on Wald Chi-Square

3.2 EVALUATION OF SAFETY

Dr. Kim Witzmann reviewed the safety of iNO in detail. The reader is referred to Dr. Witzmann's review for information regarding the adverse event profile

4 FINDINGS IN SUBGROUPS AND SPECIAL POPULATIONS

The Applicant conducted subgroup analyses by birthweight, PMA, sex, and mother's race using Fisher's exact test in the ITT population. The ITT population includes only the randomized sibling as part of the ITT population in the evaluation, and did not include the cogestational siblings. In their analyses, they found a higher percentage of successes in the iNO group compared to placebo, but none were significant with the exception of mother's race stratification of 'black'. They calculated that 56% in iNO were successes compared to only 33%. Because co-gestational siblings were included in the study and they were not randomized, I conducted subgroup analyses using GEE model with exchangeable correlation matrix, with treatment by covariate interaction. One may argue if this is necessary given that co-gestational siblings may fall in different subgroups. Even if that is the case, applying this model is better than excluding the co-gestational siblings in the analysis.

The results from subgroup analyses are summarized in Table 12. There were no significant treatment-by-subgroup interactions detected in all subgroups on the primary endpoint with the exception of age at study entry, and possibly on mother's race. For infants 7 – 14 days at enrollment, iNO showed better results than placebo in reducing the risk of being alive without CLD at 36 weeks GA. In contrast, there was no difference between treatments in the group of infants between the ages of 15 and 21 days. In the mother's race stratification, there is no numerical difference in treatment effect on "white", while there is a treatment difference in the race stratification of "black" and "others." Breslow-Day test of homogeneity of the odds ratio is significant when controlling for mother's race.

In other subgroups, I find that there is numerically higher percentage of successes in the iNO group compared to placebo in each of the subgroup.

Table 12: Subgroup Analyses of the Primary Outcome– Survival without BPD

	INO	Placebo	OR (95% CI)	p-value	Interaction Test*
Overall	**N=294**	**N=288**			
	129 (44%)	**105 (36%)**	**1.45 (1.03, 2.04)**	**0.0329****	
Sex					
Male	61/155 (39%)	54/162 (33%)	1.4 (0.8, 2.2)		0.6259
Female	68/139 (49%)	51/126 (40%)	1.5 (0.9, 2.5)		
Maternal Race					
White	55/161 (34%)	48/141 (34%)	1.1 (0.7, 1.8)		0.1175†
Black	41/75 (55%)	30/89 (34%)	2.6 (1.4, 4.9)†		
Others	27/46 (59%)	24/52 (46%)	1.6 (0.7, 3.7)		
Gestational Age					
< 26 Weeks	71/171 (42%)	65/174 (37%)	1.3 (0.9, 2.1)		0.6399
≥ 26 Weeks	58/123 (47%)	40/114 (35%)	1.7 (0.98, 2.9)		
Age of Entry					
7 – 14 days	55/112 (49%)	31/115 (27%)	2.9 (1.6, 5.3)		0.0035
15 – 21 days	74/182 (41%)	74/173 (43%)	0.97 (0.6, 1.5)		
Birth Weight					
500-749 g	68/149 (46%)	61/159 (38%)	1.52 (0.95, 2.4)		0.8441
750-999 g	47/115 (41%)	37/104 (36%)	1.30 (0.8, 2.3)		
1,000-1,250 g	14/30 (47%)	7/25 (28%)	2.06 (0.6, 6.6)		

* Generalized score tests for Type III contrasts for GEE model
** adjusted by weight strata and site
† no adjustments (model not converging when adjusted by weight and site)

5 SUMMARY AND CONCLUSIONS

5.1 STATISTICAL ISSUES AND COLLECTIVE EVIDENCE

After careful review of the application, several issues were identified. Some where addressed by the Applicant in their response to the Information Request and some where addressed by reanalyzing the data. Three statistical issues were identified that warranted further consideration when interpreting the results. The three issues are:

1. Data Quality
2. Multiplicity – Application of different statistical tests to the primary endpoint
3. Interim Analysis – level of significance adjustment

Given that this study was originally investigator-initiated, the protocol may have been written to satisfy a research hypothesis and not intended for marketing application. The standard we used in research may be not as stringent, albeit should be, compared to regulatory standard.

In summary, the difference between treatment groups is about 8.9% in favor of iNO (95%CI: 0.8%, 17.1%). This implies that for every 11 patients treated with iNO, one patient will survive without BPD, with a 95% confidence interval that ranges from 1:6 to 1:125. In addition, for infants 7 – 14 days at enrollment, iNO showed better results than placebo in reducing the risk of being alive without CLD at 36 weeks GA. In contrast, there was no difference between treatments in the group of infants between the ages of 15 and 21 days. In the mother's race stratification, there is no numerical difference in treatment effect on "white", while treatment difference favoring iNO is seen in non-whites.

In general, data quality and possible multiplicity remains potential issues. Therefore, the evidence from this study is insufficient to support the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study.

5.2 CONCLUSIONS AND RECOMMENDATIONS

This statistical review evaluates the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study. From a statistical perspective, because of the issues surrounding the design and analyses of the BALLR1 study (discussed above), it is difficult to conclude with much confidence about the mortality benefit of iNO in preterm infants at high risk for developing bronchopulmonary dysplasia at 36 weeks' post-menstrual age. Therefore I conclude that there is insufficient evidence to support the efficacy of iNOmax (nitric oxide) for the treatment of chronic lung disease for premature infant in the BALLR1 study. (b) (4) .

6 LABELING

On July 25, 2010, the Applicant submitted a supplemental NDA (Serial No. 11) to revise the current approved label with the following:

Current Pediatric Use Section of the INOmax Label:

Nitric Oxide for inhalation has been studied in a neonatal population (up to 14 days of age). No information about its effectiveness in other age populations is available.

Proposed Pediatric Use Section of the INOmax Label:

(b) (4)

FDA Proposed Text

The safety and efficacy of INOmax for the prevention of chronic lung disease [bronchopulmonary dysplasia, (BPD)] in neonates ≤ 34 weeks gestational age requiring respiratory support has been studied in three large multi-center double-blind, placebo-controlled clinical trials in a total of (b) (4) preterm infants. Of these, (b) (4) received placebo, and (b) (4) received inhaled nitric oxide at doses ranging from 5-20 ppm, for treatment periods of 7-24 days duration. The primary endpoint for these studies was alive and without BPD at 36 weeks postmenstrual age. The need for supplemental oxygen at 36 weeks PMA served as a surrogate endpoint for the presence of BPD. Overall, efficacy for the prevention of bronchopulmonary dysplasia in preterm infants was not established. There were no meaningful differences between treatment groups with regard to deaths, methemoglobin levels, or adverse events commonly observed in premature infants, including intraventricular hemorrhage, patent ductus arteriosus, pulmonary hemorrhage, and retinopathy of prematurity.

The use of INOmax for prevention of BPD in preterm neonates ≤ 34 weeks gestational age is not indicated.

7 APPENDIX

Appendix 1:

Ballard Study Discussion
Excerpt from, Interim Analysis to DSMB (March 2004)

IV. Primary and Secondary Efficacy Outcomes

This section of the report is based on the 344 infants who completed the study treatment, were discharged from the hospital, or died, and the case report forms (CRFs) were entered in the database. CRFs were completed and entered into the database for 29 of the 30 deaths to date. The only CRF not available is for the infant who died on 2/27/2004.

1. Design, Interim Looks and Primary Outcome Testing

According to the revised plan for efficacy interim analyses (submitted in October, 2001), it was decided that two interim analyses would be done with the first look at 20% of outcome data. We had originally estimated a rate of twins at 17% and at the revised plan we had estimated a rate of 7% since the observed rate was low (3.5%) at that time of enrollment. Thus, at that time it was estimated that a total of 690 infants of 644 mothers would be required in order to detect an increase in survival without CLD at 36 weeks post-menstrual age from 50% to 62% with 85% power, while controlling for Type I two-sided error of 0.05 and 2 interim analyses using the O'Brien-Fleming stopping rule and stopping for rejection of either H_0 or H_1.

Since that time, the software to calculate the sample size has been slightly improved (East 3.0, 2003). The calculation yielding 644 mothers assumed that the two looks would be equally spaced, at 33% and 67% outcome information available. From a theoretical calculation perspective, if the looks occur slightly earlier (at the planned 20% and 50%), slightly less information is available, spending a bit less of the error rate, resulting in a slightly smaller sample size. Since the calculation in 2001 was conservative, and at worst case, over-estimated the sample size, the sample size of 644 mothers remained as the proposed sample size. With the software slightly improved now to allow for specifying information time at unequal spacing, we have a requirement of 615 mothers for looks at 20% and 50% outcome.

Another aspect to consider is that all calculations were done at the 85% power to detect the expected difference from 50% to 62%. One may consider amending the sample size to allow for only 80% power, and not 85% power to detect this difference. The choice of 80% power is a standard choice taken in many studies, and may be considered here. However, since clearly 80% power will decrease the sample size required, this will change the relative spacing of the looks as well.

Note that any of these possibly proposed small changes in design arise from theoretical considerations (except for being aware of the accrual rates) and not as a result of any outcome data.

We detail in the next page the analysis plan and results based on the October 2001 premise, the slightly revised premise allowing for the unequal spacing, and a revision allowing for 80% power, instead of 85% power, acknowledging the actual information time of the previous interim analysis and the current interim analysis.

Summary of Enrolled Infants by Multiple Birth Status

	Number of Infants	Number of Independent infants
Singleton	247	247
One of twins enrolled	59	59
One of triplet enrolled	1	1
Both twins enrolled	30	15
Two of triplet enrolled	2	1
Two from quintuplet enrolled	2	1
Entire triplet enrolled	3	1
Total	344	325

1

Ballard Study Discussion
Excerpt from, Interim Analysis to DSMB (March 2004)

In order to correctly assess the information time for analysis, we need to calculate the number of independent observations. The Table in the pervious page shows this calculation. In each true multiplicity case (e.g., both twins enrolled), one was chosen at random to be included in the primary outcome analysis.

The Table below gives the design parameters under the various options. All scenarios are based on an overall α=.05, two sided testing, testing for both superiority and futility, equal randomization to both treatments, detecting differences in the primary outcome of survival without CLD from 50% to 62%, O'Brien-Fleming stopping rules and a Lan-DeMets use function.

Summary of Design Scenarios and Stopping Rules

Design features	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Spacing of looks from 100% information time	33%, 67%	20%, 50%	20%, 50%	25%, 60%
Power	85%	85%	80%	80%
Maximum accrual of mothers	626	615	539	544
Sample size at 1st look	209	123	108	136
Boundary to reject H_0 at 1st look	+3.710	+4.877	+4.877	±4.333
p-value to reject H_0 at 1st look	p<.00021	p<.000001	p<.000001	p<.000015
Boundary to reject H_1 at 1st look	±.024	0	0	±.007
p-value to reject H_1 at 1st look	p>.98	N/A	N/A	p>.99
β spent	.004			.002
Sample size at 2nd look	417	307	270	327
Boundary to reject H_0 at 2nd look	±2.511	±2.963	±2.963	±2.669
p-value to reject H_0 at 2nd look	p<.012	p<.0031	p<.0031	p<.0076
Boundary to reject H_1 at 2nd look	±.8971	±0.275	±.346	±.707
p-value to reject H_1 at 2nd look	p>.37	p>.78	p>.73	p>.48
β spent	.058	.024	.040	.067
Final boundary	±1.966	±1.957	±1.953	±1.957
p-value to reject H_0	.049	.050	.051	.050
Total sample size of infants (based on 7% twins)	673	661	580	585

At the first interim report, 142 infants of 136 mothers were included, and the chi-squared statistic had p=.134. Thus, no stopping rule was met for any of the scenarios. At this look, there are 344 infants of 325 mothers. Table 14a gives the results of survival without CLD for the 325 infants selected at random (for the multiplicity cases) of the 344 infants. Table 14b gives the results for all 344 infants, for completeness of data presentation, but the statistics are based on Table 14a. The chi-squared statistics is 1.715, corresponding to a test statistic for the stopping rules of 1.3096, and p=.19. Employing the Lan-DeMets adjustment to adjust the stopping rules in the table above to the actual number of outcome points observed in the analysis, gives stopping rules ranging between ±2.662 (Scenario 3) to ±2.899 (Scenario 1) to reject H_0, and stopping rules of ±.347 (Scenario 1) and ±.707 (Scenario 3) to reject H_1. The test statistic at 1.3096 is not close to any of these boundaries, indicating that no stopping rules are met. The percentage information included in this interim analysis, based on design Scenarios 1 and 2 at this look are 52%, and are 60% for design Scenarios 3 and 4.

Appendix 2: List of tables that were omitted or modified and the reasoning (BALLR1 Study):

Table 4: Changed from "List of All Patients Who Did Not Tolerate Study Gas Re-initiation (Intent-to-treat Patients)" to "List of All Patients Who Did Not Complete the Study (Safety Population). The latter table was considered more relevant than the former.

Table 10B: Primary Outcome Stratified By Severity of Disease Based On RSS – Respiratory Severity Score at Baseline (Intent- To-Treat Population). Table was omitted due to the limited amount of data available for RSS.

Tables 15A and 15B: Secondary Outcome: Change in Airway Resistance. These tables were omitted due to the fact that Airway Resistance was not collected on the CRF, and was only collected on supplementary forms at several sites.

Tables 16I-K: These were adverse event tables based on the designation of "Serious". Unfortunately, this information was not collected.

Table 18: Incidence of Elevated Nitrogen Dioxide Concentrations (Safety Population). Nitrogen Dioxide was not collected on the case report forms.

Appendix 3: Demographic and Baseline Characteristics by Treatment Received (Safety Population)

Variable		iNO	Placebo)	Total
Postmenstrual age (weeks)	N	295	289	584
	Mean (SD)	25.7 (1.43)	25.7 (1.48)	25.7 (1.45)
	Median	25.6	25.3	25.4
	Range	(22.9, 30.4)	(22.0, 31.0)	(22.0, 31.0)
	< 26 weeks; n (%)	171 (58)	174 (60)	345 (59)
	≥ 26 weeks; n (%)	124 (42)	115 (40)	239 (41)
Sex; n (%)	Male	156 (53)	162 (56)	318 (54)
	Female	139 (47)	127 (44)	266 (46)
Weight (g)	N	295	289	584
	Mean (SD)	765.3 (160.71)	759.1 (154.63)	762.2 (157.63)
	Median	743.0	726.0	739.5
	Range	(500.0, 1247.0)	(500.0, 1227.0)	(500.0, 1247.0)
Prenatal history; n (%)	Vaginal-spontaneous	108 (37)	107 (37)	215 (37)
	Vaginal-unknown	3 (1)	8 (3)	11 (2)
	Caesarean	184 (62)	174 (60)	358 (61)
Respiratory severity score	n	102	96	198
	Mean (SD)	4.2 (2.03)	4.1 (2.16)	4.1 (2.09)
	Median	3.5	3.5	3.5
	Range	(1.2, 10.8)	(1.6, 16.1)	(1.2, 16.1)
Respiratory severity score strata; n (%)	< 3.5	48 (16)	48 (17)	96 (16)
	3.5 - < 10	51 (17)	46 (16)	97 (17)
	≥ 10	3 (1)	2 (1)	5 (1)
	Missing	193 (65)	193 (67)	386 (66)
Apgar scores, 1 minute	n	290	287	577
	Mean (SD)	4.1 (2.22)	4.1 (2.34)	4.1 (2.28)
	Median	4.0	4.0	4.0
	Range	(0.0, 8.0)	(0.0, 9.0)	(0.0, 9.0)

Variable		iNO (N = 295)	Placebo (N = 289)	Total (N = 584)
Apgar scores, 5 minute	n	290	286	576
	Mean (SD)	6.4 (1.93)	6.3 (1.94)	6.4 (1.93)
	Median	7.0	7.0	7.0
	Range	(0.0. 9.0)	(0.0, 10.0)	(0.0, 10.0)
Apgar scores, 10 minute	n	116	110	226
	Mean (SD)	6.5 (1.76)	6.4 (1.68)	6.5 (1.72)
	Median	7.0	7.0	7.0
	Range	(1.0, 10.0)	(1.0, 10.0)	(1.0, 10.0)
Baseline/entry ventilation; n (%)	CMV-P	183 (62)	169 (58)	352 (60)
	CMV-V	19 (6)	18 (6)	37 (6)
	HFJV	6 (2)	2 (1)	8 (1)
	HFOV	59 (20)	72 (25)	131 (22)
	NP IMV	2 (1)	5 (2)	7 (1)
	NP CPAP	25 (8)	18 (6)	43 (7)
	Unknown	1 (0)	5 (2)	6 (1)
Multiple births; n (%)	Single	211 (72)	214 (74)	425 (73)
	Twins	63 (21)	62 (21)	125 (21)
	Triplets	10 (3)	6 (2)	16 (3)
	Quadruplets	0 (0)	1 (0)	1 (0)
	Missing	11 (4)	6 (2)	17 (3)

SD = standard deviation.
Note: Numbers may not total to expected sample size due to missing data.
Sources: Section 14.1.2, Table 8

Source: BALLR1 Study Report, Table 9

Appendix 4: Maternal Demographic and Baseline Characteristics by Treatment Received (Safety Population)

Variable		iNO (N = 295)	Placebo (N = 289)	Total (N = 584)
Maternal Race; n (%)	White	161 (55)	142 (49)	303 (52)
	Black	74 (25)	90 (31)	164 (28)
	Other	47 (16)	52 (18)	99 (17)
	Unknown	13 (4)	5 (2)	18 (3)
Age of Mother (Years)	n	289	285	574
	Mean (SD)	26.9 (6.83)	27.7 (6.25)	27.3 (6.55)
	Median	27.0	28.0	27.0
	Range	(14.0, 47.0)	(15.0, 42.0)	(14.0, 47.0)

SD = standard deviation
Note: Numbers may not total to expected sample size due to missing data.
Source: Section 14.1.2, Table 8

Source: BALLR1 Study Report, Table 10

This is a representation of an electronic record that was signed electronically and this page is the manifestation of the electronic signature.

/s/

JOAN K BUENCONSEJO
11/23/2010

THOMAS J PERMUTT
11/24/2010
concur