Sliced Testis Assay Prevalidation: Baseline/Positive Control Study

WA 3-10, Task 7

REVISED OVERALL TASK DRAFT FINAL REPORT*

Authors (Overall Task Report):

Jerry D. Johnson, Ph.D.
Diplomate, A.B.T.
Work Assignment Leader
Battelle Memorial Institute

Paul I. Feder, Ph.D.
Senior Statistician
EDSP Data Coordination Center Manager
Battelle Memorial Institute

Performing Laboratory:

Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201-2693

Sponsor:

Environmental Protection Agency
Endocrine Disruptor Screening Program
Washington, D.C.

Sponsor's Representative:

David P. Houchens, Ph.D.
EDSP Program Manager
Battelle Memorial Institute

Study Initiation Date:

January 20, 2004

In-Life Performance Dates:

January 29 – March 17, 2004

Task Initiation Date:

September, 2004

Draft Final Report Date: March 21, 2005

*The original overall draft report included data from one laboratory that used $\log_{10}$ in its statistical analysis rather than natural logs.

REVISED OVERALL TASK DRAFT FINAL REPORT*

Title: Sliced Testis Assay Prevalidation: Baseline/Positive Control Study

Overall Task Report Authors/Lab:

Jerry D. Johnson, Ph.D.
Diplomate, A.B.T.
Work Assignment Leader
Battelle Memorial Institute

Paul I. Feder, Ph.D.
Senior Statistician
EDSP Data Coordination Center Manager
Battelle Memorial Institute

Lead/Participating Lab Study Directors:

Carol S. Sloan
Research Triangle Institute International (Lead Laboratory)
Research Triangle Park, NC

Joyce Durnford, Ph.D.
Battelle Memorial Institute
Columbus, OH

Paul M. Lezberg, B.A.
Toxikon Corporation
Bedford, MA

Gary A. Piazza, Ph.D.
Southern Research Institute
Birmingham, AL

Christopher J. Bowman, Ph.D.
WIL Research Laboratories, Inc.
Ashland, OH

Sponsor: Environmental Protection Agency
Endocrine Disruptor Screening Program
Washington, D.C.

Sponsor's Representative: David P. Houchens, Ph.D.
EDSP Program Manager
Battelle Memorial Institute

*The original overall draft report included data from one laboratory that used $\log_{10}$ in its statistical analysis rather than natural logs.

Authors:

Signature Date
Jerry D. Johnson, Ph.D.
Diplomat, A.B.T.
Work Assignment Leader

Signature Date
Paul I. Feder, Ph.D.
Senior Statistician
EDSP Data Coordination Center Manager

Approved:

Signature Date
David P. Houchens, Ph.D.
EDSP Program Manager

GOOD LABORATORY PRACTICES COMPLIANCE STATEMENT

This study was conducted in compliance with EPA's GLP Regulations 40 CFR Part 160 and 40 CFR Part 792. This study was conducted according to the study QAPP and each participating laboratory's study protocols and Standard Operating Procedures (except where noted in appendices). The data presented accurately reflect the results of the study.

Jerry D, Johnson, Ph.D., D.A.B.T
Work Assignment Leader

Quality Assurance Statement

This study was inspected by the Quality Assurance Unit and reports were submitted to the Work Assignment Leader (WAL) and management as follows:

Phase Inspected	Inspection Date	Date Reported to the WAL and WAL Management
Interlaboratory statistics report	7/8/04	7/8/04
Overall task report	11/1/04	11/1/04
Revised overall task report	3/18/05	3/18/05

Quality Assurance Unit **Date**

Table of Contents

List of Figures

List of Tables

List of Appendices

EXECUTIVE SUMMARY

The purpose of Work Assignment (WA) 3-10 was to perform a series of prevalidation studies using the sliced testis assay to develop positive and cytotoxicant control data, demonstrate the relevance of the assay using reference chemicals, and obtain sufficient data on intra- and interlaboratory variability to permit the design of the validation studies. The present report addresses the Task 7 study results for WA 3-10. The objectives of Task 7 were to:

- Collect baseline and positive and cytotoxicant control data for the sliced testis assay
- Evaluate aminoglutethimide (AG) as a positive control for the assay
- Evaluate ethane dimethane sulfonate (EDS) as a cytotoxicant control for the assay
- Determine baseline and positive control inter- and intralaboratory variability.

A total of five laboratories, one lead laboratory and four other participating laboratories, took part in conducting this task.

Briefly, the sliced testis assay procedure used the right testis from male Sprague-Dawley rats, 11-15 weeks old. A testis was sliced to yield fragments weighing 50-100 mg, which were placed into individual tubes that contained 95% O_2 / 5% CO_2 freshly gassed modified medium-199 without phenol red. Fragments were allowed to equilibrate in control media prior to being tested in control media, vehicle control media, or test chemical-treated media with or without human chorionic gonadotropin (hCG). Media samples were collected after time 0 (baseline), 1, 2, 3, and 4 hours of incubation. The baseline sample and composite sample (0-4 hours) from each fragment were analyzed for testosterone and lactate dehydrogenase (LDH). Three replicates of the design summarized in the following table were conducted by each laboratory.

Sample Type	hCG	Number of Incubations (Runs)	Testis Fragments
Media-Vehicle control	no	3	1 - 3
Media-Vehicle control	yes	3	4 - 6
Media control	yes	3	7 - 9
Media + AG (10 µM)	yes	3	10 - 12
Media + AG (100 µM)	yes	3	13 - 15
Media + AG (1000 µM)	yes	3	16 - 18
EDS (Cytotoxicant Control)	yes	3	19 - 21

Testosterone

Overall mean testosterone baseline (0 hr) concentration was 0.05 ng/mg with an average within laboratory sd of ±0.02 ng/mg and a between laboratories sd of ±0.02 ng/mg.

For unstimulated fragments, the overall mean testosterone composite (0-4 hr) concentration was 0.69 ng/mg with an average within laboratory sd of ±0.25 ng/mg and a

between laboratories sd of ±0.24 ng/mg. The hourly average testosterone concentration produced for unstimulated fragments was 0.16 ng testosterone/mg fragment/hr.

For media-vehicle (M-V)-Control hCG-stimulated fragments, the overall mean testosterone composite concentration was 4.28 ng/mg with an average within laboratory sd of ±0.97 ng/mg and a between laboratories sd of ±2.13 ng/mg. The hourly average testosterone concentration production for hCG stimulated fragments was 1.06 ng testosterone/mg fragment/hr. There was a 7.5 ±1.5 fold increase in testosterone production for the stimulated relative to the unstimulated fragments when determined for all replicates and laboratories. Stimulated Media-Control fragments produced similar results.

For AG-treated fragments, a concentration-dependent decrease in testosterone concentration was observed. The overall mean testosterone composite concentrations were 2.47, 0.71, and 0.19 ng/mg for the 10, 100, and 1000 µM AG groups, respectively. The overall average within laboratory sd values were 0.65, 0.19, and 0.06 ng/mg and the overall between laboratory sd values were 1.61, 0.45, and 0.06 ng/mg for the 10, 100, and 1000 µM AG groups, respectively. There was a -45.0 ±4.9, -84.4 ±2.0, and -96.2 ±0.6 percent decrease in the testosterone concentration at 10, 100, and 1000 µM AG when determined for all replicates and laboratories.

For EDS-treated fragments, testosterone concentrations were decreased. The overall mean testosterone composite concentration was 1.06 ng/mg with an average within laboratory sd of 0.35 ng/mg and a between laboratories sd of 0.62 ng/mg. There was a -76.6 ±2.0 percent decrease in testosterone concentration at 1000 µM EDS when determined for all replicates and laboratories.

Testosterone Interlaboratory Statistical Analysis

All laboratories except Lab B used a natural-logarithmic transformation. Lab B used a base 10 logarithmic transformation. Therefore, for the interlaboratory analysis the reported values for each endpoint for Lab B were multiplied by ln(10)=2.3026 in order to transform them to the natural-logarithmic scale, to be in conformance with the other laboratories. It should also be noted that results from Lab B were not transformed in the Lab B intralaboratory analysis reports, since they are self consistent.

Comparisons among the laboratories using the Q-statistics for homogeneity of laboratory results indicated that there was significant heterogeneity among the laboratory results for comparisons between the (M-V)-control hCG-unstimulated and -stimulated groups.

Significant differences existed between the 10, 100 and 1000 µM AG groups and the M-V +hCG group across all laboratories as well as within laboratories. Significant linear components of trend occurred among the M-V +hCG group and the three graded dose AG groups across all laboratories as well as within each laboratory. Significant within laboratory quadratic components of trend among the M-V +hCG group and the three graded dose AG groups occurred across all laboratories, as well as within Labs A, B, and C. Differences between the M-V +hCG group and both the M-V -hCG and the Cytotoxicant EDS +hCG control groups

were significant across all laboratories as well as within each laboratory. There were no significant differences between the M-V +hCG group and the M +hCG group across all laboratories or within any laboratory.

LDH

Overall mean LDH baseline (0 hr) concentration was 1.29 mU/mg with an average within laboratory sd of ±0.28 mU/mg and a between laboratories sd of ±0.30 mU/mg.

For unstimulated fragments, the overall mean LDH composite (0-4 hr) concentration was 8.4 mU/mg with an average within laboratory sd of ±1.3 mU/mg and a between laboratories sd of ±2.0 mU/mg.

For M-V-Control hCG-stimulated fragments, the overall mean LDH composite concentration was 8.0 mU/mg with an average within laboratory sd of ±0.8 mU/mg and a between laboratories sd of ±1.7 mU/mg. These results were similar to that obtained without stimulation. Stimulated Media-Control fragments produced similar results.

For AG-treated fragments, LDH was not concentration-dependent. The overall mean LDH composite concentrations were 7.9, 7.9, and 7.8 mU/mg for the 10, 100, and 1000 µM AG groups, respectively. These values were similar to the M-V-control (stimulated and unstimulated) and M-control values.

For EDS-treated fragments, the overall mean LDH composite concentration was 8.5 mU/mg with an average within laboratory sd of 1.3 mU/mg and a between laboratories sd of 1.6 mU/mg. Treatment with EDS, at a concentration of 1000 µM and under the conditions of the present study, resulted in similar LDH concentrations as measured with the control groups.

LDH Interlaboratory Statistical Analysis

The Q-statistic for consistency among laboratory results show that there was no significant heterogeneity across the laboratory results. No significant differences existed between any of the AG groups and the M-V +hCG groups, either across all laboratories or within each laboratory. A significant linear component of trend occurred among the M-V +hCG control group and the three graded dose AG groups within Lab D. The difference between the M-V +hCG and M-V -hCG groups was significant across all laboratories combined but not within any individual laboratory. Nor was there any significant difference between the M-V +hCG control group and EDS-treated group across all laboratories combined, although it was significant for Lab E.

Conclusions

The sliced testis assay was conducted by five laboratories to determine testosterone and LDH concentration baseline levels and effects of treatment with positive (AG) and cytotoxicant (EDS) controls. Measures of variability were determined for control and treatment groups within and across laboratories. AG, at concentrations ranging from 10 to 1000 µM, produced a

statistically significant concentration-dependent decrease in testosterone concentration across all laboratories and within all laboratories. AG treatment had no effect on LDH. EDS, at a concentration of 1000 µM, significantly decreased testosterone concentration, but not LDH, across all laboratories and within all laboratories. AG would be an effective positive control test substance for the assay; however, EDS was equivocal as a cytotoxicant based on LDH as the endpoint for assessing cytoxicity.

1.0 INTRODUCTION AND BACKGROUND

The Food Quality Protection Act of 1996 requires the EPA to develop and implement a screening program using valid tests for determining the potential in humans for estrogenic effects from pesticides. EPA proposed a two-tiered screening program in a Federal Register notice in 1998 (63 FR 71542-71568, Dec. 28, 1998) that covered not only pesticides but also commercial chemicals subject to regulation under the Toxic Substances Control Act (TSCA; 15 USC 2601) and environmental and drinking water contaminants. One of the assays recommended for validation and consideration for inclusion in the screening program is an *in vitro* assay for steroidogenesis. A detailed review paper (DRP) reviewed the different types of steroidogenesis assays. On the basis of recommendations in the DRP, the *in vitro* sliced testes steroidogenesis assay was selected as the most promising screening tool for identifying substances with steroidogenic-altering activity. Despite a long history of use, the sliced tissue assay had not been optimized. Optimization of the sliced testes assay was performed under Work Assignment (WA) 2-27. The resulting protocol needed to be validated prior to its use as the basis for a test guideline in the Endocrine Disruptor Screening Program (EDSP).

The purpose of WA 3-10 was to perform a series of prevalidation studies to develop control chemical data, demonstrate the relevance of the assay, and obtain sufficient data on intra- and interlaboratory variability to permit the design of the validation studies. WA 3-10 was divided into several tasks of which three involved conducting studies. Task 5 was a study of several cytotoxic chemicals that was conducted to determine the performance of the assay to detect Leydig cell toxicity and to choose a reference cytotoxicant for the assay system. Task 7 was a baseline positive control study, run without reference chemicals and with the reference cytotoxicant to determine whether the assay was performing correctly and to generate data on within-lab (intralaboratory) and lab-to-lab (interlaboratory) variability. Task 8 consisted of multichemical studies conducted in the lead laboratory to demonstrate the sensitivity and relevance of the assay.

The present report addresses the Task 7 study results. The objectives of this task were to:

- Collect additional baseline and positive control data for the sliced testis assay
- Evaluate aminoglutethimide (AG) as a positive control for the assay
- Evaluate ethane dimethane sulfonate (EDS) as a cytotoxicant control
- Determine baseline and positive control inter- and intralaboratory variability.

A total of five laboratories, one lead laboratory and four other participating laboratories, took part in conducting this task. Although the work assignment called for three, rather than four participating laboratories, a fourth laboratory was added for the following reasons:

- Based on an analysis of the confidence interval factors for the coefficient of variation across laboratories, a dramatic improvement in power was obtained by using three laboratories and a further improvement was observed by using four laboratories, whereas additional gain required using 8 or 20 laboratories (see Minced Testes Steroidogenesis Assay Sensitivity Analysis to Provide a Rationale for the Number of

Laboratories to be Selected; Appendix A).

- The validation of this assay may require as many as six laboratories. Thus, by including a fourth laboratory in the prevalidation studies, the Environmental Protection Agency (EPA) would have more laboratories with experience for participating in the validation studies.

- A fourth laboratory provided a contingency lab in case something unforeseen occurred, thereby providing assurance that the prevalidation work was completed at the minimum level needed to meet the objectives of the work assignment.

The present report consists of an overall Task report and individual lead and other participating laboratory reports. The overall Task report was written by Battelle and includes salient information from the individual laboratory study reports, Battelle's interlaboratory statistical analysis, and accounts of laboratory issues that impacted study conduct. The individual laboratory reports were written by the lead and participating laboratories and includes their intralaboratory statistical variability analysis.

2.0 METHODS AND MATERIALS

2.1 Chemistry

Chemistry support for this study was provided by the Chemical Repository (WA 4-2). The chemistry support included chemical procurement, formulation method development and validation, analysis method development and validation, purity, stability, formulation preparation, and formulation analysis. All of these activities and the results are described in detail for each test chemical in chemical-specific reports that are included in Appendix B. The salient chemistry information from these reports is summarized in Table 2.1-1. The chemistry activities performed by the lead and participating laboratories involved diluting the stock solution (prepared, analyzed, and shipped by the Chemical Repository) to the appropriate concentrations for testing. The procedures for receipt, dilution of the stock, and return shipment of residual test chemical formulations are described in the individual laboratory reports.

The target concentration of the stock formulation prepared by the Chemical Repository and the dilutions that were prepared by the laboratories, as well as the final target concentrations tested in the incubation mixture are summarized in Table 2.1-2.

Table 2.1-1. Salient Test Chemical Information

Chemical	CAS No.	Purity Mfr/Battelle	Stock Formulation Vehicle	Stock Formulation Stability	Analysis Method
Aminoglutethimide (AG)	125-84-8	99% / 99%	DMSO (100 %)	>39 days	Gas chromatography with flame ionization detection
Ethane dimethane sulfonate (EDS)	4672-49-5	98% / 100%	DMSO (100%)	>31 days	Gas chromatography with flame ionization detection

Table 2.1-2 Target Stock and Dilution Formulation and Final Incubation Concentrations

Test Chemical	Target Stock Concentration in DMSO (mg/mL)[a]	Target Stock Concentration in modified M-199 (mg/mL)[b]	Final Target Concentration in the Incubation Media (µM)
AG (high)	23.2	0.232	1000
AG (mid)	2.32	0.0232	100
AG (low)	0.232	0.00232	10
EDS	21.8	0.218	1000

a For AG, 1:10 and 1:100 dilutions of the high AG-DMSO stock were used to prepare the mid and low AG-DMSO stocks, respectively.

b Each AG-DMSO stock formulation was diluted 1:100 in modified M-199.

2.2 Laboratory and Assay Conduct Information

Five laboratories were involved in the conduct of this task. One laboratory was designated as the lead laboratory (RTI) and four additional laboratories were included as participating laboratories. The lead laboratory provided assay conduct training to the participating laboratory staff and the results of this training are reported in WA 3-10, Task 6. Summary information about the laboratories and study conduct milestones are included in Table 2.2-1.

Table 2.2-1. Laboratory and Assay Conduct Information

Laboratory	Location	Study Director	Start Date (Protocol Sig.)	In-Life Assay Performance Date
Research Triangle Institute (RTI)	Research Triangle Park, NC	Carol S. Sloan	Jan. 20, 2004	Feb. 19 - Mar. 2, 2004
Battelle	Columbus, OH	Dr. Joyce Durnford	Jan. 27, 2004	Jan. 29 - Mar. 17, 2004
Southern Research Institute (SRI)	Birmingham, AL	Dr. Gary A. Piazza	Jan. 29, 2004	Feb. 2 -Mar. 4, 2004
Toxikon	Bedford, MA	Paul M. Lezberg	Jan. 26, 2004	Feb.18 - Mar. 2, 2004
WIL Research Laboratories	Ashland, OH	Dr. Christopher J. Bowman	Jan. 23, 2004	Feb. 4 - 13, 2004

2.3 Assay Procedure

The assay procedures of the lead and participating laboratories are described in the individual reports from each laboratory. The procedure is illustrated in Figure 2.3-1.

The assay was conducted at a time during the morning that enabled the technician(s) to isolate and initiate fragment equilibration and incubation for a replicate experiment before approximately noon on a given day. Male Sprague-Dawley rats, 11-15 weeks old, with testes weighing greater than 1000 mg, were used. The rats were euthanized using carbon dioxide and the right testis was surgically isolated and removed. The testis was weighed (to the nearest 0.1 g) and transferred to a petri dish where the tunica albicans was removed. Next, the testis was sliced, first longitudinally, and then in cross sections to yield fragments weighing 50-100 mg. The fragments were weighed (to the nearest 0.0001 g) and the weights recorded. The target time from testis removal to the time of slicing was approximately 30 minutes. Each fragment was placed into individual tightly capped 9 mL test tubes that contained 2.5 mL of 95% O_2 / 5% CO_2 freshly gassed media (room temperature; modified medium-199 without phenol red; pH approximately 7.4).

After all fragments were collected and transferred to media-containing test tubes, the test tubes were then placed on a shaker (175 $\pm$ 5 rpm) that was located inside an incubator (36°C $\pm$ 0.5°C). After 30 minutes (Equilibration Phase), the test tubes were removed from the shaker/incubator and centrifuged (800 x g for 5 min, 2-8°C). The supernatant was poured off and discarded (equilibration wash). Fresh media (2.5 mL) was added to the test tubes and centrifuged (800 x g for 5 min, 2-8°C). The supernatant was collected and saved for analysis (Baseline Collection Phase Sample at Time 0 hr).

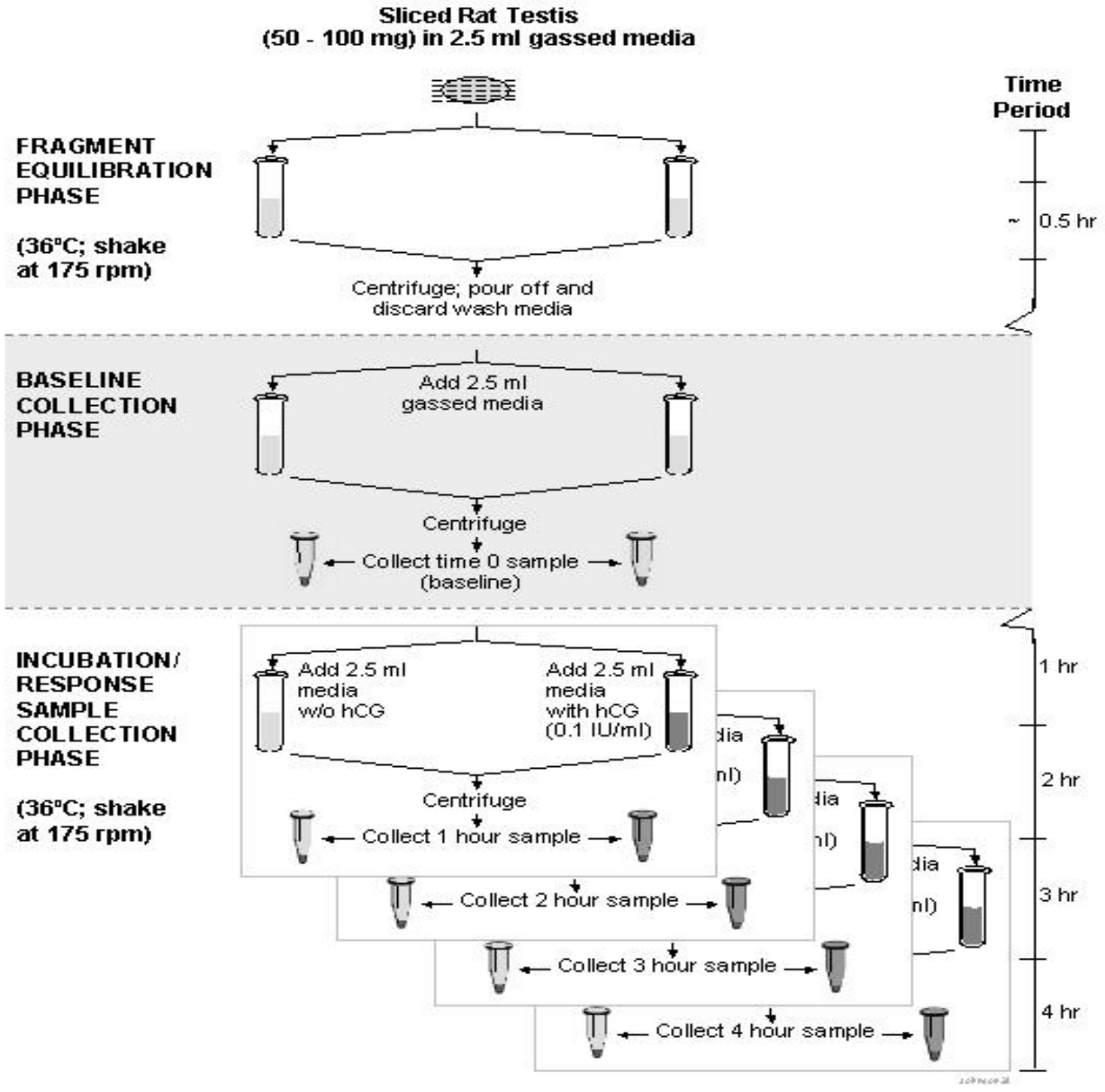

Figure 2.3-1. Technical Flow Illustration of the Sliced Testis Assay

Fragments were distributed among the various control and treatment groups (see Study Design section). Group specific media was used to replenish media that was collected at all specified time points for the various control and treatment groups. Human chorionic gonadotropin (hCG) (0.1 U/mL, final concentation) was used to stimulate steroidogenesis. Media types included media control with hCG, media-vehicle control with and without hCG, media with positive control (AG) at each of three concentration levels and hCG, and media with a cytotoxic negative control (EDS) at one concentration level and hCG. Fresh media (2.5 mL) was added, according to group, and the test tubes were returned to the shaker/incubator, which began the Incubation Response Sample Collection Phase. After 1 hour, the test tubes were removed from shaker/incubator, centrifuged (800 x g for 5 min, 2-8°C), and the supernatant (media) collected (Sample at Time 1 hr). This procedure was repeated for collection of supernatant (media) samples at 2, 3, and 4 hours (Samples at Time 2 hr, Time 3 hr, and Time 4 hr). After collection of the baseline and all hourly samples for a given fragment, a composite sample was prepared by combining a 0.5 mL aliquot from the baseline and each hourly sample for a given fragment into a single container (final volume 2.5 mL).

The original baseline sample and composite sample from each fragment were analyzed for testosterone (radioimmunoassay [RIA] method) and lactate dehydrogenase (LDH) (spectrophotometric method). For testosterone determinations, media samples were analyzed undiluted or diluted 1:10 in duplicate. Testosterone samples were stored at -70 to -80°C prior to analysis. For LDH determinations, analysis of the baseline and composite samples were performed within the same day that the sliced testis assay was conducted. LDH samples were stored at room temperature and protected from the light prior to analysis.

2.4 Animal Information

Salient information related to the animals and their care and use for each of the laboratories is summarized in Table 2.4-1. All laboratories used male Sprague-Dawley rats (Sprague-Dawley Derived Outbred Albino rat [Crl:CD(SD)IGS BR]) from Charles River Laboratories, Inc. (Raleigh, NC). All laboratories provided the animals *ad libitum* access to feed (certified rodent diet No. 5002, except for SRI that used Teklad 2018 Certified Global Diet from Harlan, Madison, WI) and water (municipal supply) and used a 12 hour light:dark cycle. Quarantine periods were at least seven days. All laboratories housed the animals for this study in separate study rooms with no other animals or test chemicals.

2.5 Study Design

The experimental design for the Baseline/Positive Control study for prevalidation is summarized in Table 2.5-1.

Table 2.4-1. Animal and Husbandry Information by Laboratory

Laboratory	Animal Receipt Date	Animal Age Upon Receipt	Animal Use Age	Housing	Environmental Conditions
RTI	Not reported	10 weeks	11-12 weeks	Individual, solid-bottom polycarbonate cages	Temp Range: 64 - 79° F RH Range: 30 - 70%
Battelle	38014	10 weeks	12-15 weeks	Individual, solid-bottom polycarbonate cages	Temp Range: 69 - 75° F RH Range: 35 - 65%
SRI	February 3, 2004	11 weeks	14-15 weeks	Individual, solid-bottom polycarbonate cages	Temp Range: 64 - 79° F RH Range: 36-87%
Toxikon	Not reported	10 weeks	12-14 weeks	Individual, solid-bottom polycarbonate cages	Temp Range: 64 - 79° F RH Range: 30 - 70%
WIL	January 28, 2004	10 weeks	11-12 weeks	Individual, stainless-steel, wire-mesh cages	Temp Range: 67.9 - 69.5° F RH Range: 42.9 - 60.4%

Table 2.5-1. Summary of the Experimenal Design

Sample Type	hCG	Number of Incubations (Runs)	Testis Fragment(s)
Media-Vehicle control	no	3	1 - 3
Media-Vehicle control	yes	3	4 - 6
Media control	yes	3	7 - 9
Media + AG (10 µM)	yes	3	10 - 12
Media + AG (100 µM)	yes	3	13 - 15
Media + AG (1000 µM)	yes	3	16 - 18
EDS (Cytotoxicant Control)	yes	3	19 - 21

The information presented in the table represents one replicate of the experiment. Three replicate experiments were conducted. The overall study used nine rats and three rats/replicate study; one right testes/rat; three testes total/replicate study; seven fragments/testis. A block design was used for distribution of testis fragments to reduce variability. The seven fragments obtained from each testis were divided among the seven test conditions.

The sampling time points (5) from the media were 0 (after a 30 min. equilibration) and 1, 2, 3, and 4 hours post-equilibration. A 0.5 ml aliquot was taken from samples taken at 0, 1, 2, 3, and 4 hours and combined to prepare a single composite sample for each fragment. The 0 hr and composite samples were then analyzed for testosterone in duplicate and LDH in singlet.

2.6 Statistical Methods

Statistical analysis was divided into intralaboratory and interlaboratory components. Each laboratory carried out the intralaboratory analyses individually, based on a common analysis plan developed by the Battelle Data Coordination Center (DCC). The analysis plan is summarized below. In addition, each individual laboratory report contains its intralaboratory analysis. The interlaboratory analysis, performed by Battelle, is presented in Appendix C.

It is important to note that one laboratory used $\log_{10}$ for its statistical analysis. The present revised overall report converted the data to natural log in order to make interlaboratory comparisons.

2.6.1 Intralaboratory Statistical Analysis

The intralaboratory statistical analysis plan, as provided by Battelle to the lead and participating laboratories, is summarized below.

All analyses were carried out on the fragment weight adjusted concentrations for both testosterone (ng/mg) and LDH (mU/mg). The data set consisted of analyzing the results of seven test groups: four control groups and three graded aminoglutethimide (AG) groups. These were referenced in the analyses as:

- M-V + hCG — Stimulated media-vehicle control
- M-V - hCG — Unstimulated media-vehicle control
- M + hCG — Stimulated media control (no vehicle)
- M-V + hCG EDS — Stimulated media-vehicle cytoxicant control

- AG 10 + hCG — 10 μM AG group
- AG 100 + hCG — 100 μM AG group
- AG 1000 + hCG — 1000 μM AG group.

The 0 to 4 hour composite determinations were adjusted by subtracting the corresponding baseline values. Analyses were based on the natural logarithms of the baseline-adjusted values, except for Toxicon which used $\log_{10}$. Preliminary summaries and graphical displays were prepared based on summary statistics of these logarithmic differences, by test group. These summaries provided initial indications of response trends and any variance heterogeneity.

The intralaboratory statistical analysis was divided into two parts. Part 1 was a comparison among the four control groups. Part 2 was a comparison among the (M-V + hCG) group and the three graded dose AG groups. In each part the (M-V + hCG) group was the comparison standard.

For each analysis part, mixed effects analysis of variance models were fitted to the data. Random effects were replicates, testes (or equivalently animals) within replicates, and fragments within testes. Fixed effects were the four classification groups (within each part), treated as classification factors. For each part a model was fitted to the data assuming heterogeneous fragment-to-fragment variances among test groups and an alternative model was fitted assuming common fragment-to-fragment variances among test groups. The separate variance and common variance models were compared by a likelihood ratio test. Comparisons among test groups were carried out using the separate variance or the common variance models, depending on the results of the likelihood ratio tests. In both the Part 1 and Part 2 analyses least squares means and associated standard errors were reported for each test group. Pairwise differences of each group with the M-V + hCG group were reported, with associated standard errors and significance levels.

In the Part 2 analysis linear and quadratic contrasts in the graded AG concentration groups were reported. The contrasts treated the (M-V +hCG) group as 0 mM (actually 1 mM, since it is mapped into 0 on the log scale).

Predicted and residual values were calculated based on the fits to each of the models. The residuals were plotted versus replicate and versus test group to assess the homogeneity of variability across groups and to identify possible outlying fragments. A normal probability plot of the residuals was prepared to assess conformity to model assumptions and to identify possible outlying fragments.

For the Part 2 analysis predicted values were plotted versus ordered test groups to assess the response trend with increasing AG concentration levels.

For each testis within each replicate a logarithmic mean value was calculated across the seven test groups. Standard errors of the logarithmic averages were estimated based on the variance components estimated within each model fit. These standard errors were used to calculate control limits to identify possible outlying testes or heterogeneous replicates based on those testis or replicate averages that exceeded three standard errors.

All of the laboratories assumed homogeneous variance across groups to model the fragment-to-fragment variation within groups with the following exceptions: (1) Battelle used a heterogeneous variance model for the Part 2 analysis for testosterone concentrations and for the Part 1 analysis for LDH concentrations; (2) Toxikon used heterogeneous variance for Parts 1 and 2 for LDH; and (3) RTI used heterogeneous variance for Part 2 for LDH. Battelle and Toxikon Laboratories used different normalizing factors than those specified in the analysis plan to estimate the linear and quadratic trend contrasts. In order to combine the trend contrasts across all the laboratories their contrast values and associated standard errors were properly scaled to conform to the analysis plan.

2.6.2 Interlaboratory Statistical Analysis

All laboratories except Toxikon used a natural-logarithmic transformation. Toxikon used a base 10 logarithmic transformation. Therefore, for the interlaboratory analysis the reported values for each endpoint for Toxikon were multiplied by ln(10)=2.3026 in order to transform them to the natural-logarithmic scale, to be in conformance with the other laboratories. It should also be noted that results from Toxikon were not transformed in the Toxikon intralaboratory analysis reports, since they are self consistent. The complete interlaboratory statistical analysis report is included in Appendix C.

The objectives of the interlaboratory statistical analysis were to:

a. Determine the extent of agreement among laboratories with respect to:

   - Differences between each of the graded doses of AG and the stimulated media-vehicle control,
   - Slope of the dose trend, and
   - Differences between the stimulated media-vehicle control group and each of the three alternative control groups; and

b. Estimate the coefficients of variation among laboratories for each of the above endpoints. A supplemental analysis on the coefficients of variation across laboratories for testosterone concentrations was also performed and this supplemental report is included in Appendix D.

Similar statistical analyses were carried out for the testosterone concentration and for the LDH concentration responses. For each of testosterone and LDH concentrations separate statistical analyses were carried out for each of the seven endpoints discussed above for the intralaboratory analysis section.

A weighted one-way analysis of variance statistic Q was calculated to evaluate the consistency of analysis results across the five laboratories. Q is defined as

$$Q= \Sigma[(Y_i - Y_w)^2/S_i^2]$$

where $Y_w = \Sigma[(1/S_i^2)Y_i /(\Sigma 1/S_i^2)]$ is a weighted average over the five laboratories, and Y_i and S_i are the effect and the within laboratory standard error reported by the i^{th} laboratory. Under the null hypothesis of homogeneous mean effects across laboratories, Q is distributed approximately as a chi square with 4 (number of laboratories -1) degrees of freedom (DerSimonian and Laird, 1986). Normal probability plots of the reported estimates across laboratories were prepared to assess the extent of homogeneity across laboratories and to determine whether departures from homogeneity represented random variation across laboratories or whether there were systematic differences among laboratories or outlying laboratories.

For each endpoint a one-way mixed effects analysis of variance model with heterogeneous variances among the participating laboratories was fitted to the mean responses using weights incorporating within laboratory variances. The random effect was laboratory. The within-laboratory variances were the squares of the standard errors reported by each laboratory. The analysis of variance fit provided an estimated weighted average effect across all laboratories and its associated standard error. The degrees of freedom associated with the overall effect was calculated (based on Satterthwaite's approximation) as:

$$2*[((1/K)*\Sigma\ (S_L^2 + S_i^2))^2]/[(var(S_L^2) +(2/K^2)*\ \Sigma\ (S_i^4/df_i))]$$

where S_L^2 is the random laboratory to laboratory variance, S_i^2 and df_i are the reported within laboratory variance and degrees of freedom for the i^{th} laboratory, $var(S_L^2)$ is the variance of S_L^2, and K is the number of laboratories (Hartung and Makambi, 2001).

For all comparisons statistical significance was set at the 0.05 level.

For each endpoint, the estimated overall average effect and its associated standard error (incorporating laboratory to laboratory variation) and degrees of freedom were used to construct a 95% confidence interval. The individual effects and associated 95% confidence intervals (based on the within laboratory standard error) for each laboratory were also determined. These were plotted side-by-side to provide a graphical comparison among the laboratories.

To describe the variability among the laboratories relative to the average effect value, coefficients of variation (CV) and their associated 95% confidence intervals (CI) were calculated. The coefficient of variation is defined as the standard deviation of the effect response divided by its mean. Note that the CV pertains to the measured concentrations (and not their logarithmic transformations). Since the measurements are assumed to be approximately log normally distributed the CV may be expressed as

$$CV=([\exp(S^2)-1])^{½}$$

where S^2 is the total variance among the five laboratories. The 95% CI is based on the chi square distribution and is calculated as

$$[(\exp(df*S^2/(\chi^2_{df,\ 0.975}))- 1))^{½},\ (\exp(df*S^2/(\chi^2_{df,\ 0.025}))- 1)^{½}]$$

where df is the estimated degree of freedom among the five laboratories, indicated above.

2.7 On-Site Laboratory Monitoring

Dr. Jerry D. Johnson, the work assignment leader (WAL), visited three of the five laboratories during the conduct of Task 7. He visited RTI, SRI, and Battelle but was unable to visit WIL and Toxikon. (The group at Battelle that conducted the assay reported to a different

division director than Dr. Johnson.) His visits occurred when the laboratories were conducting their third replicate so as to ascertain the level of performance the laboratories had achieved after the experience they gained by completing their first and second replicates. So as to minimize his influence on laboratory methods and procedures, Dr. Johnson remained an impartial and silent observer, taking notes and making observations. Relevant and salient information collected during the site visits are included in the results section of this report.

2.8 Good Laboratory Practices (GLP) Compliance

The laboratories conducted this task in conformance to applicable laws and regulations. Specific regulatory requirements included the current EPA Good Laboratory Practices regulations as set forth in 40 CFR Part 792 and 40 CFR Part 160; and when sections of the regulations were not performed by the laboratory or under the direction of the laboratory, then these activities were called out in the report. In addition, a Quality Assurance Project Plan (QAPP) was prepared by Battelle in conjunction with RTI. The QAPP was submitted to each of the laboratories prior to study start for usage by their respective Quality Assurance Units. The QAPP is included in Appendix E.

2.9 Archives

All task documents generated by the individual laboratories are retained at the respective laboratories' archival facilities. Chemistry activity documents (Chemical Repository), individual laboratory spreadsheets and intralaboratory statistical analysis reports, as well as the interlaboratory statistical analysis report are stored at Battelle's archival facilities and/or the EDSP DCC.

3.0 RESULTS

The remainder of the report uses codes for laboratory identification so that the assay results can be the focus of the data analysis and interpretation, rather than the laboratory where the work was conducted. The individual laboratory reports with their appendices are included in Appendices F, G, H, I, and J for Laboratories A, B, C, D, and E, respectively.

3.1 Testosterone

3.1.1 Testosterone Baseline Sample (0 Hr) Results

Testosterone concentrations (ng/mg) measured at the 0 hour sampling time point (Baseline) are summarized in Table 3.1-1. Since the baseline sample was obtained from the media prior to initiating fragment treatment, the baseline values for all groups were used to determine individual laboratory and overall mean baseline values.

Laboratory mean baseline values ranged from 0.03 to 0.07 ng/mg with CV percentages ranging from 28.3 to 63.4 percent. The CV percentages approximated 30 percent for four of five labs. The one lab with a CV percentage value of 63 percent had consistently low baseline values for replicates 1 and 2 but much higher baseline values for replicate 3. No explanation for this

increase was uncovered. The overall mean baseline value was 0.05 ng/mg with an average within laboratory sd of ±0.02 ng/mg and a between laboratories sd[1] of ±0.02 ng/mg. The overall CV percentage was 40 percent.

Table 3.1-1. Baseline (0 Hr) Testosterone Concentrations (ng/mg) by Treatment and Laboratory*

Treatment Group		Laboratories				
	Replicate	Lab A	Lab B	Lab C	Lab D	Lab E
M-V Control - hCG	1	0.02	0.08	0.06	0.05	0.02
	2	0.02	0.08	0.04	0.03	0.06
	3	0.08	0.06	0.06	0.02	0.04
M-V Control +hCG	1	0.02	0.10	0.09	0.06	0.03
	2	0.02	0.09	0.05	0.03	0.06
	3	0.06	0.04	0.05	0.02	0.04
Media Control +hCG	1	0.01	0.09	0.09	0.04	0.03
	2	0.03	0.09	0.05	0.04	0.05
	3	0.06	0.05	0.05	0.02	0.04
AG @ 10 uM + hCG	1	0.02	0.09	0.08	0.05	0.03
	2	0.02	0.10	0.05	0.04	0.06
	3	0.06	0.04	0.04	0.02	0.03
AG @ 100 uM + hCG	1	0.02	0.10	0.08	0.05	0.02
	2	0.02	0.09	0.04	0.04	0.04
	3	0.07	0.05	0.05	0.03	0.04
AG @ 1000 uM + hCG	1	0.02	0.09	0.09	0.04	0.03
	2	0.02	0.08	0.05	0.04	0.05
	3	0.06	0.05	0.05	0.03	0.03
EDS +hCG	1	0.02	0.09	0.08	0.04	0.03
	2	0.02	0.06	0.07	0.04	0.05
	3	0.06	0.05	0.04	0.03	0.03
Within Laboratory Mean		0.03	0.07	0.06	0.04	0.04
sd		0.02	0.02	0.02	0.01	0.01
CV%		63.4	28.3	29.8	30.9	32.1
Overall Mean		0.05				
Average Within sd		0.02				
Overall Between sd		0.02				
*Each laboratory cell is an average of three fragments.						

3.1.2 Testosterone Composite Sample (4 Hr) Results

Testosterone concentrations (ng/mg) measured at the 4 hour sampling time point (Composite) are summarized in Tables 3.1-2 and 3.1-3. Each treatment group was evaluated separately since the testosterone concentration produced during the incubation period was treatment-dependent. The overall laboratory results are illustrated in Figure 3.1-1.

[1]The between laboratories sd is the sd between the laboratory sample averages.

Table 3.1-2. Composite (4 Hr) Testosterone Concentrations (ng/mg) for Media Groups by Laboratory*

Treatment Group		Laboratories				
	Replicate	Lab A	Lab B	Lab C	Lab D	Lab E
M-V Control - hCG	1	0.54	0.84	0.76	1.16	0.24
	2	0.48	1.56	0.82	0.68	0.57
	3	0.71	0.68	0.65	0.39	0.30
Within Laboratory Mean		0.58	1.03	0.74	0.74	0.37
sd		0.12	0.47	0.09	0.39	0.18
CV%		20.7	45.7	11.6	52.3	47.5
Overall Mean		0.69				
Average Within sd		0.25				
Overall Between sd		0.24				
M-V Control +hCG	1	3.52	3.33	7.57	2.66	3.17
	2	3.48	5.61	6.87	3.58	2.29
	3	5.14	3.32	9.23	1.71	2.74
Within Laboratory Mean		4.05	4.09	7.89	2.65	2.73
sd		0.95	1.32	1.21	0.94	0.44
CV%		23.4	32.3	15.4	35.3	16.1
Overall Mean		4.28				
Average Within sd		0.97				
Overall Between sd		2.13				
Media Control +hCG	1	2.89	3.26	8.45	1.27	2.11
	2	4.25	6.17	7.55	3.13	2.53
	3	5.35	3.81	8.46	1.86	2.14
Within Laboratory Mean		4.16	4.41	8.15	2.09	2.26
sd		1.23	1.55	0.52	0.95	0.23
CV%		29.6	35.0	6.4	45.6	10.4
Overall Mean		4.22				
Average Within sd		0.90				
Overall Between sd		2.44				
*Each laboratory cell is an average of three fragments.						

Table 3.1-3. Composite (4 Hr) Testosterone Concentrations (ng/mg) for Treatment Groups by Laboratory*

Treatment Group		Laboratories				
	Replicate	Lab A	Lab B	Lab C	Lab D	Lab E
AG @ 10 uM + hCG	1	2.63	2.55	5.60	0.70	1.93
	2	1.31	3.26	3.93	1.47	1.61
	3	2.33	1.96	5.92	0.84	1.05
Within Laboratory Mean		2.09	2.59	5.15	1.00	1.53
sd		0.69	0.65	1.07	0.41	0.45
CV%		33.1	25.1	20.7	40.9	29.1
Overall Mean		2.47				
Average Within sd		0.65				
Overall Between sd		1.61				
AG @ 100 uM + hCG	1	0.34	0.95	1.87	0.29	0.27
	2	0.42	0.78	0.89	0.52	0.32
	3	0.66	0.93	1.48	0.45	0.46
Within Laboratory Mean		0.47	0.89	1.41	0.42	0.35
sd		0.17	0.09	0.49	0.12	0.10
CV%		35.2	10.5	34.9	28.1	28.1
Overall Mean		0.71				
Average Within sd		0.19				
Overall Between sd		0.45				
AG @ 1000 uM + hCG	1	0.09	0.32	0.31	0.16	0.11
	2	0.13	0.27	0.18	0.16	0.21
	3	0.22	0.25	0.20	0.10	0.13
Within Laboratory Mean		0.15	0.28	0.23	0.14	0.15
sd		0.07	0.04	0.07	0.03	0.05
CV%		45.4	12.9	30.4	24.7	35.3
Overall Mean		0.19				
Average Within sd		0.05				
Overall Between sd		0.06				
EDS +hCG	1	0.70	0.77	2.20	0.40	0.40
	2	0.91	1.15	1.42	1.27	0.59
	3	1.20	1.33	2.65	0.52	0.42
Within Laboratory Mean		0.94	1.08	2.09	0.73	0.47
sd		0.25	0.29	0.62	0.47	0.10
CV%		26.8	26.4	29.8	64.6	22.2
Overall Mean		1.06				
Average Within sd		0.35				
Overall Between sd		0.62				

*Each laboratory cell is an average of three fragments.

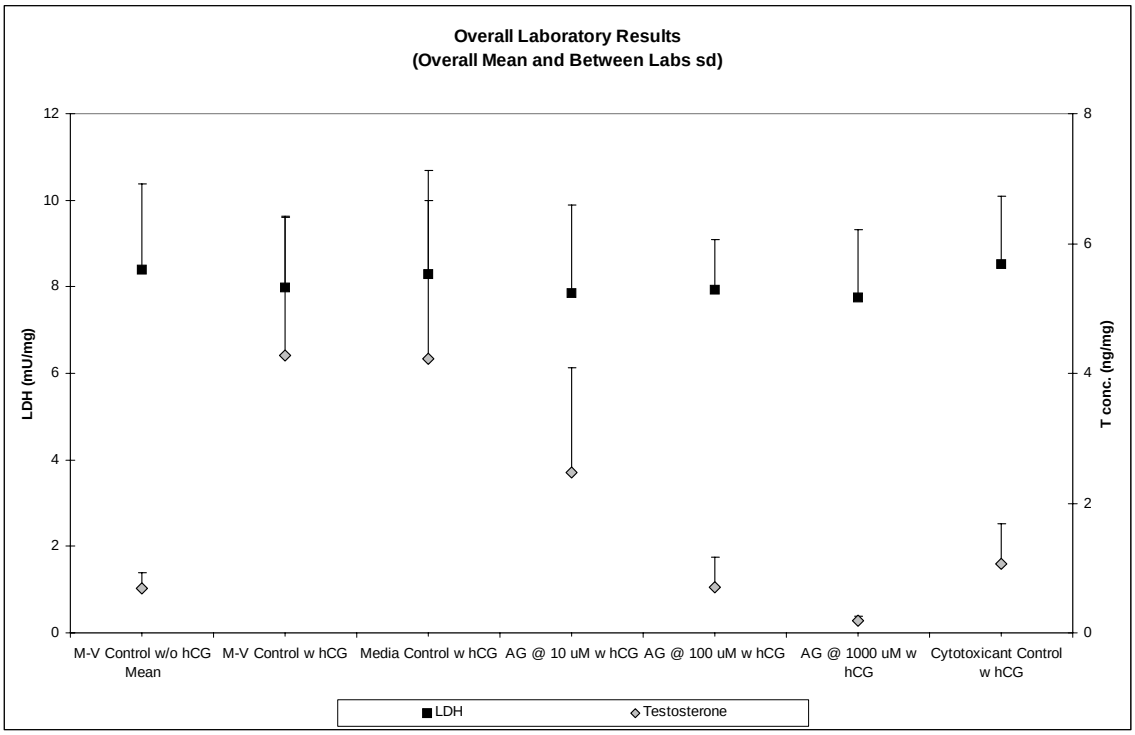

Figure 3.1-1. Overall Mean and Between Laboratories sd for LDH and Testosterone (4 Hr Composite Values)

Laboratory M-V-control w/o hCG (unstimulated) mean composite values ranged from 0.37 to 1.03 ng/mg (Table 3.1-2). CV percentages[2] approximated 50 percent for three laboratories, whereas two of the laboratories had values of 21 (Lab A) and 12 (Lab C) percent. The overall mean composite value was 0.69 ng/mg with an average within laboratory sd of ±0.25 ng/mg and a between laboratories sd of ±0.24 ng/mg[3]. The hourly average testosterone concentration produced for unstimulated fragments was 0.16 ng testosterone/mg fragment/hr (0.69 - 0.05 ng/mg divided by 4 hours).

Laboratory M-V-control w/hCG (stimulated) mean composite values ranged from 2.65 to 7.89 ng/mg (Table 3.1-2). CV percentages varied widely from 15 to 35 percent. The overall mean composite value was 4.28 ng/mg with an average within laboratory sd of ±0.97 ng/mg and a between laboratories sd of ±2.13 ng/mg. The hourly average testosterone concentration

[2]The CV% shown in Tables 3.1-2 and 3.1-3 corresponds to the variation between the three replicate mean concentrations, each of which is an average of three testis concentrations.

[3]The between laboratories sd is the sd between the within laboratory means.

production for stimulated fragments was 1.06 ng testosterone/mg fragment/hr (4.28 - 0.05 ng/mg divided by 4 hours). This represents a 6.6-fold increase/hour in testosterone production for the stimulated relative to the unstimulated fragments.

Laboratory media-control w/hCG (stimulated) mean composite values ranged from 2.09 to 8.15 ng/mg (Table 3.1-2). CV percentages varied widely from 6 to 46 percent. The overall mean composite value was 4.22 ng/mg with an average within laboratory sd of ±0.90 ng/mg and a between laboratories sd of ±2.44 ng/mg. The hourly average testosterone concentration production for stimulated fragments was 1.04 ng testosterone/mg fragment/hr (4.22 - 0.05 ng/mg divided by 4 hours). This represents a 6.5-fold increase/hour in testosterone production for the stimulated relative to the unstimulated fragment. M-control w/hCG (stimulated) mean composite values resulted in similar testosterone concentrations to those produced by M-V-control w/hCG, thereby suggesting that the vehicle (DMSO, 1 percent, v/v) did not affect fragment response (Table 3.1-2).

AG produced a concentration-dependent decrease in testosterone concentration in stimulated fragments (w/hCG) and this finding was consistent for all laboratories (Table 3.1-3). Laboratory AG w/hCG (stimulated) mean composite values ranged from 1.00 to 5.15, 0.35 to 1.41, and 0.14 to 0.28 ng/mg for the 10, 100, and 1000 µM AG groups, respectively. CV percentages varied from 11 to 45 percent. The overall mean composite values were 2.47, 0.71, and 0.19 ng/mg for the 10, 100, and 1000 µM AG groups, respectively. The average within laboratory sd values were 0.65, 0.19, and 0.06 ng/mg and the overall between laboratory sd values were 1.61, 0.45, and 0.06 ng/mg for the 10, 100, and 1000 µM AG groups, respectively. Based on the overall mean composite values for the M-V-control w/hCG and the AG groups, a relative percentage decrease in testosterone concentration following 10, 100, and 1000 µM AG was 42.3, 83.4, and 95.6 percent, respectively.

EDS decreased the testosterone concentration in stimulated fragments (w/hCG) and this finding was consistent for all laboratories (Table 3.1-3). Laboratory EDS w/hCG (stimulated) mean composite values ranged from 0.47 to 2.09 ng/mg. CV percentages were similar for most labs at approximately 25 percent, except for one laboratory (Lab D) that had a value of 65 percent. The overall mean composite value was 1.06 ng/mg with an average within laboratory sd of 0.35 ng/mg and a between laboratories sd of 0.62 ng/mg. Based on the overall mean composite values for the M-V-control w/hCG and EDS group, a relative percentage decrease in testosterone concentration following 1000 µM EDS was 75.2 percent.

3.2 LDH Analysis

3.2.1 LDH Baseline Sample (0 Hr) Results

LDH concentrations (mU/mg) measured at the 0 hour sampling time point (Baseline) are summarized in Table 3.2-1. Since the baseline sample was obtained from the media prior to initiating fragment treatment, the baseline values for all groups were used to determine individual laboratory and overall mean baseline values. Laboratory mean baseline values ranged from 0.84 to 1.61 mU/mg with CV percentages ranging from 9.2 to 39.5 percent. The

overall mean baseline value was 1.29 mU/mg with an average within laboratory sd of ±0.28 mU/mg and a between laboratories sd[1] of ±0.30 mU/mg.

Table 3.2-1. Baseline (0 Hr) LDH Concentrations (mU/mg) by Treatment and Laboratory*

Treatment Group		Laboratories				
	Replicate	Lab A	Lab B	Lab C	Lab D	Lab E
M-V Control - hCG	1	0.86	1.02	1.45	2.15	0.72
	2	1.09	1.10	1.26	1.04	0.74
	3	1.29	1.70	1.86	1.55	0.86
M-V Control +hCG	1	0.88	1.02	2.07	1.91	0.79
	2	1.07	2.11	1.19	1.38	0.88
	3	1.25	1.48	1.35	1.63	0.90
Media Control +hCG	1	0.98	1.16	1.54	1.45	0.72
	2	1.70	3.36	1.20	1.67	0.87
	3	1.20	1.59	1.45	1.67	0.88
AG @ 10 uM + hCG	1	0.89	1.03	1.66	1.75	0.94
	2	1.29	1.88	1.18	1.49	0.81
	3	1.29	1.09	1.22	1.74	0.85
AG @ 100 uM + hCG	1	1.06	0.95	1.61	1.64	0.71
	2	1.50	1.28	1.17	1.40	0.89
	3	1.49	1.32	1.53	1.65	0.80
AG @ 1000 uM + hCG	1	0.80	1.05	1.64	1.45	0.76
	2	1.29	1.28	1.24	1.52	0.98
	3	0.95	1.45	1.17	1.51	0.82
EDS +hCG	1	1.41	0.98	1.47	1.93	0.88
	2	1.13	1.08	1.82	1.43	0.93
	3	1.18	1.27	0.93	1.85	0.87
Within Laboratory Mean		1.17	1.39	1.43	1.61	0.84
sd		0.24	0.55	0.28	0.24	0.08
CV%		20.2	39.5	19.6	14.8	9.2
Overall Mean		1.29				
Overall Within sd		0.28				
Overall Between sd		0.30				

*Each laboratory cell is an average of three fragments.

[1]The between laboratories sd is the sd between the laboratory sample averages.

3.2.2 LDH Composite Sample (4 Hr) Results

LDH concentrations (mU/mg) measured at the 4 hour sampling time point (Composite) are summarized in Tables 3.2-2 and 3.2-3. Each group was evaluated separately since the LDH concentration produced during the incubation period was treatment-dependent. The overall laboratory results are illustrated in Figure 3.1-1.

Table 3.2-2. Composite (4 Hr) LDH Concentrations (mU/mg) for Media Groups by Laboratory*

Treatment Group		Laboratories				
	Replicate	Lab A	Lab B	Lab C	Lab D	Lab E
M-V Control - hCG	1	5.35	6.29	8.99	11.21	8.27
	2	4.98	8.22	9.53	8.19	7.41
	3	5.35	12.10	11.64	10.46	8.06
Within Laboratory Mean		5.23	8.87	10.05	9.95	7.91
sd		0.21	2.96	1.40	1.57	0.45
CV%		4.1	33.4	13.9	15.8	5.7
Overall Mean		8.40				
Average Within sd		1.32				
Overall Between sd		1.98				
M-V Control +hCG	1	5.42	6.82	9.68	10.51	8.22
	2	5.44	9.82	8.58	8.02	7.07
	3	5.38	7.26	9.62	10.22	7.64
Within Laboratory Mean		5.41	7.97	9.29	9.58	7.64
sd		0.03	1.62	0.62	1.36	0.58
CV%		0.6	20.3	6.7	14.2	7.5
Overall Mean		7.98				
Average Within sd		0.84				
Overall Between sd		1.66				
Media Control +hCG	1	6.07	8.89	9.89	11.63	8.10
	2	5.87	10.33	9.15	8.25	7.83
	3	4.76	10.02	9.88	9.58	8.08
Within Laboratory Mean		5.57	9.75	9.64	9.82	8.00
sd		0.71	0.76	0.42	1.70	0.15
CV%		12.7	7.8	4.4	17.3	1.9
Overall Mean		8.56				
Average Within sd		0.75				
Overall Between sd		1.83				
*Each laboratory cell is an average of three fragments.						

Table 3.2-3. Composite (4 Hr) LDH Concentrations (mU/mg) for Treatment Groups by Laboratory*

Treatment Group		Laboratories				
	Replicate	Lab A	Lab B	Lab C	Lab D	Lab E
AG @ 10 uM + hCG	1	5.77	6.85	10.03	10.99	9.55
	2	5.31	5.93	8.53	8.60	6.16
	3	4.49	6.51	10.31	10.25	8.62
Within Laboratory Mean		5.19	6.43	9.62	9.95	8.11
sd		0.65	0.47	0.96	1.22	1.75
CV%		12.5	7.2	9.9	12.3	21.6
Overall Mean		7.86				
Average Within sd		1.01				
Overall Between sd		2.04				
AG @ 100 uM + hCG	1	6.99	7.72	9.64	9.61	8.76
	2	5.93	5.13	8.75	7.37	8.13
	3	5.99	9.02	9.00	9.13	7.87
Within Laboratory Mean		6.30	7.29	9.13	8.70	8.25
sd		0.60	1.98	0.46	1.18	0.46
CV%		9.4	27.2	5.0	13.6	5.5
Overall Mean		7.94				
Average Within sd		0.93				
Overall Between sd		1.14				
AG @ 1000 uM + hCG	1	5.12	6.94	9.33	9.82	9.53
	2	5.51	4.76	9.12	7.84	7.87
	3	5.42	9.32	9.37	8.25	8.06
Within Laboratory Mean		5.35	7.01	9.27	8.64	8.49
sd		0.20	2.28	0.13	1.05	0.91
CV%		3.8	32.6	1.4	12.1	10.7
Overall Mean		7.75				
Average Within sd		0.91				
Overall Between sd		1.58				
EDS +hCG	1	7.80	7.33	10.33	11.47	10.44
	2	5.34	5.12	9.86	7.95	8.50
	3	6.52	9.00	9.59	9.62	9.13
Within Laboratory Mean		6.55	7.15	9.93	9.68	9.36
sd		1.23	1.95	0.37	1.76	0.99
CV%		18.8	27.2	3.8	18.2	10.6
Overall Mean		8.53				
Average Within sd		1.26				
Overall Between sd		1.56				

*Each laboratory cell is an average of three fragments.

Laboratory M-V-control w/o hCG (unstimulated) mean composite values ranged from 5.2 to 10.1 mU/mg (Table 3.2-2). CV percentages[2] were all below 20 percent, except for one laboratory (Lab B) that had a value of 33 percent. The overall mean composite value was 8.4 mU/mg with an average within laboratory sd of ±1.3 mU/mg and a between laboratories sd[3] of ±2.0 mU/mg.

Laboratory M-V-control w/hCG (stimulated) mean composite values ranged from 5.4 to 9.6 mU/mg (Table 3.2-2). CV percentages were all at or below 20 percent. The overall mean composite value was 8.0 mU/mg with an average within laboratory sd of ±0.8 mU/mg and a between laboratories sd of ±1.7 mU/mg. These results were similar to those obtained without stimulation.

Laboratory media-control w/hCG (stimulated) mean composite values ranged from 5.6 to 9.8 mU/mg (Table 3.2-2). CV percentages were all below 20 percent. The overall mean composite value was 8.6 mU/mg with an average within laboratories sd of ±0.8 mU/mg and a between laboratories sd of ±1.8 mU/mg. M-control w/hCG (stimulated) mean composite values resulted in similar LDH concentrations to those produced by M-V-control w/hCG, thereby suggesting that the vehicle (DMSO, 1 percent, v/v) did not affect fragment response (Table 3.2-2).

Treatment with AG, at three concentration levels ranging from 10 to 1000 µM, resulted in similar LDH concentrations for all AG concentrations and all laboratories (Table 3.2-3). Laboratory AG w/hCG (stimulated) mean composite values ranged from 5.2 to 10.0, 6.3 to 9.1, and 5.4 to 9.3 mU/mg for the 10, 100, and 1000 µM AG groups, respectively. CV percentages for a given AG concentration within laboratories were generally less than 20 percent, although one laboratory per treatment group had a value that was greater than 20 percent, (Lab E @ 10 µM, 22%; Lab B @ 100 µM, 27%; Lab B @ 1000 µM, 33%). The overall mean composite values were 7.9, 7.9, and 7.8 mU/mg for the 10, 100, and 1000 µM AG groups, respectively. These values were similar to the M-V-control (stimulated and unstimulated) and M-control values.

EDS, the cytotoxicant control, produced LDH concentrations that were similar for all laboratories (Table 3.2-3). Laboratory EDS w/hCG (stimulated) mean composite values ranged from 6.6 to 9.9 mUg/mg. CV percentages were less than 20 percent, except for one laboratory (Lab B) that had a value of 27 percent. The overall mean composite value was 8.5 mU/mg with an average within laboratory sd of 1.3 mU/mg and a between laboratories sd of 1.6 mU/mg.

---

[2] The CV% shown in Tables 3.2-2 and 3.2-3 corresponds to the variation between the three replicate mean concentrations, each of which is an average of three testis concentrations.

[3] The between laboratories sd is the sd between the within laboratory means.

3.3 Intralaboratory Statistical Analysis

Intralaboratory statistical analysis results excerpts from the individual laboratory reports are provided below. Editing of the reports for inclusion in the overall report body was required in order to provide consistent references to test groups, correctly reference figures/tables, and extract the most pertinent information. The full individual laboratory statistical analysis narratives are included with each laboratory's report (see Appendices).

The intralaboratory statistical analyses were divided into two parts. Part 1 was a comparison among the following four control groups:

- MV-Control + hCG — Stimulated media-vehicle control
- MV-Control-hCG — Unstimulated media-vehicle control
- M-Control+hCG — Stimulated media control
- EDS+hCG — Stimulated cytotoxicant control.

Part 2 was a comparison among the following groups:

- MV-Control + hCG — Stimulated media-vehicle control
- AG 10+hCG — 10 µM AG group
- AG 100+hCG — 100 µM AG group
- AG 1000+hCG — 1000 µM AG group.

3.3.1 Lab A

The full Lab A statistical analysis narrative, with graphs and tables, is included in Appendix F.

Except for the model assessing the effects of the AG treatments and the M-V+hCG control groups on LDH concentrations, the residual variances of the various treatment groups were not significantly different from one another (α=0.05). However, differences in the residual variances of the AG treatments and the M-V+hCG control group were marginally significant (P=0.065) for LDH in the AG dosage group model. Although the balanced RCB design is robust to all but extreme violations of the variance homogeneity, the investigators preferred the conservative approach of considering the AG-treatment group variances to be unequal. Accordingly, RCB modeling results for LDH for the AG-treatment groups versus the M-V +hCG control group will be reported for the heterogeneous variances case. In fact, the RCB model for LDH for the AG-treatment groups versus the M-V+hCG control group was fit under *both* the homogeneity and heterogeneity assumptions with virtually no difference in the estimated treatment effects.

For LDH, the variability due to different runs of the experiment (Rep) was much lower than the individual random variability (Residual variability) and the Testis (Replication) variability was zero. Note that for the heterogeneous variance RCB model fit to LDH for AG treatment groups versus the M-V+hCG control group, there is a separate residual variance

estimate for each treatment group; the variance of the 10 μM AG + hCG group is considerably (but not significantly) larger than that of the other three groups or that of the variability due to replication. In contrast, there was considerable variability in testosterone concentrations measured from different testes, within a given replication of the testosterone concentrations measured from different testes, within a given replication of the testosterone experiments and the Replication variance component was zero. This suggests that there is something systematically different about the individual runs of the testosterone experiments and that this difference is reflected in the degree of homogeneity of the testosterone measurements among the testicular fragments within a given run. In fact, the testes (replication) and residual variance components associated with analyses of testosterone concentration were of similar magnitude. Thus blocking on the run of the experiment and on the testis from which the fragments were obtained appears to be critical for testosterone studies but not for LDH studies. It may be prudent to investigate sources of the variability in the testosterone runs and perhaps try to control them procedurally in the laboratory.

Based on previously published studies, the differences (with 95% confidence intervals) between the simple M-V+hCG control group and each of the other control groups and the AG-treatments were all in the expected direction and of the expected magnitudes. For example, although the mean difference between the LDH concentrations is always greater than zero, the magnitude of these differences does not change systematically with the AG dose (linear contrast P=0.5503; quadratic contrast P=0.8522). This contrasts with the pattern of testosterone concentration differences, where a clear increasing trend is observed with increasing AG dosage. Although the trend could be well approximated by a simple linear process on the log scale, both the linear and the quadratic contrasts were statistically significant (linear contrast P<0.0001; quadratic contrast P=0.0085).

3.3.2 Lab B

The full Lab B statistical analysis narrative, with graphs and tables, is included in Appendix G. It is important to note that the statistical analysis results for this laboratory are expressed as $\log_{10}$ rather than natural logs.

For the LDH control group comparisons, mixed model ANOVAs were fitted to the data, assuming heterogeneous fragment-to-fragment variances among test groups, and assuming homogeneous variances. In both cases, the model included treatment group as fixed effects and random effects of replicate, animal within replicate, and fragment within testes. The heterogenous variance model fit produced a residual log likelihood of 15.6192, compared to 5.7307 for the homogeneous variance model. The chi-square value of 19.7770, with 3 degrees of freedom indicates that the homogeneous variance model should be rejected (p=0.0002). Based on the relevant inference tests from the mixed effects ANOVA, assuming heterogeneous variance, there was no significant difference among the four control groups.

For the LDH three graded aminoglutethimide (AG) dose groups versus the (M-V Control/hCG) control group, mixed model ANOVAs were fitted to the data assuming heterogeneous fragment-to-fragment variances among test groups, and assuming homogeneous variances. In both cases, the model included treatment group as fixed effect and random effects

of replicate, animal within replicate, and fragment within testes. The heterogenous variance model fit produced a residual log likelihood of 22.9204, compared to 16.1568 for the homogeneous variance model. The chi-square value of 13.5272, with 3 degrees of freedom indicates that the homogeneous variance model should be rejected (p=0.0036). Based on the relevant inference tests from the mixed effects ANOVA, assuming heterogeneous variance, there was no significant difference among the four treatment groups.

Summary statistics for log-transformed testosterone/fragment concentration by treatment were determined. The geometric mean adjusted testosterone concentration was 3.342 ng/mg for the M-V Control/hCG group compared to 0.808 ng/mg for the M-V Control w/o hCG, 0.905 for the Cytotoxicant Control/hCG, and 3.668 ng/mg for the Media Control/hCG group. The geometric mean adjusted testosterone concentration was 3.342 ng/mg for the M-V Control/hCG group compared to 2.224 ng/mg for the AG @10 μM/hCG group, 0.699 ng/mg for the AG @100 μM/hCG group, and 0.191 ng/mg for the AG @1000 μM/hCG group.

For the testosterone control group comparisons, mixed model ANOVAs were fitted to the data assuming heterogeneous fragment-to-fragment variances among test groups, and assuming homogeneous variances. In both cases, the model included treatment group as fixed effects and random effects of replicate, animal within replicate, and fragment within testes. The heterogenous variance model fit produced a residual log likelihood of 0.7710, compared to –0.2080 for the homogeneous variance model. The chi-square value of 1.9580, with 3 degrees of freedom indicates that the homogeneous variance model should not be rejected (p=0.5812).

For the relevant inference tests from the mixed effects ANOVA, assuming homogeneous variance, there was a significant difference among the control groups (p>=0.0001). The means for the M-V Control w/o hCG group (p<=0.0001) and the Cytotoxicant Control/hCG group (p<=0.0001) were significantly different from the M-V Control/hCG group.

For testosterone in the three graded aminoglutethimide (AG) dose groups versus the (M-V Control/hCG) control group, mixed model ANOVAs were fitted to the data assuming heterogeneous fragment-to-fragment variances among test groups, and assuming homogeneous variances. In both cases, the model included treatment group as fixed effects and random effects of replicate, animal within replicate, and fragment within testes. The heterogenous variance model fit produced a residual log likelihood of 5.6468, compared to 4.8219 for the homogeneous variance model. The chi-square value of 1.6498, with 3 degrees of freedom indicates that the homogeneous variance model should not be rejected (p= 0.6482).

Based on the relevant inference tests from the mixed effects ANOVA, assuming homogeneous variance, there was a significant difference among the treatment groups (p>=0.0001). The means for the AG @10 μM/hCG group (p= 0.0124), the AG @100 μM/hCG group (p<=0.0001), and the AG @1000 μM/hCG (p<=0.0001) were significantly different from the M-V Control/hCG group. Both the linear (p<=0.0001) and the quadratic (p=0.0003) contrasts were statistically significant.

3.3.3 Lab C

The full Lab C statistical analysis narrative, with graphs and tables, is included in Appendix H.

In each part, the M-V-Control+hCG group was the comparison standard. A homogenous variance type was used for both control (Part 1) and treatment (Part 2) comparisons. There were no statistically significant differences in LDH values for either Parts 1 or 2. In Part 1 for testosterone, the Least Squares Mean testosterone values for the MV-Control without hCG and Cytotoxicant Control groups were lower (statistically significant; $p \leq 0.05$) than the M-V-Control with hCG group. In Part 2, the Least Squares Mean testosterone values in the 10, 100, and 1000 µM AG with hCG groups were lower (statistically significant; $p<0.05$) than the M-V-Control with hCG group. The linear and quadratic trend estimates were statistically significant ($p<0.05$).

3.3.4 Lab D

The full Lab D statistical analysis narrative, with graphs and tables, is included in Appendix I.

The logarithmically transformed results were more symmetric, particularly in testosterone concentrations. Therefore, statistical analyses were performed on the log-transformed baseline-adjusted testosterone concentration and LDH concentration results. Also, the variabilities of the log-transformed testosterone concentrations varied among the dose groups when including all three replicates. The largest variability occurred in the AG 100 µM dose group.

Based on the likelihood ratio tests, the heterogeneous variance model was selected for the Part 2 analysis of testosterone concentrations and Part 1 analysis of LDH concentrations, while the homogeneous variance model was selected for the other two analyses. Residual plots showed that all model fits to the data were reasonable, as most residual values were systematically spread around zero. The ordered residuals against normal quantiles conformed to the distribution reference lines, and no apparent residuals departed from the reference lines. Therefore the model assumptions were correct and none of the fragments were considered as outliers.

Significant differences in testosterone concentrations existed for the 10, 100, and 1000 µM AG dose groups and the stimulated media-vehicle control ($p= 0.001$, 0.0002 and <0.0001, respectively). There was a significant negative linear dose trend ($p<0.0001$). As the dose level of AG increased, the log-transformed baseline-adjusted testosterone concentrations decreased. The estimated fragment-to-fragment variation in testosterone concentrations varied among the study groups. The AG 100 µM group had the largest variability (0.6869), while the (M-V +hCG) group and AG 1000 µM group had relatively low variabilities (0.06966 and 0.03163, respectively). The estimated replicate variation in testosterone concentrations was zero.

Based on analysis of variance results for the log-transformed baseline-adjusted testosterone concentrations among the four control groups, there were differences between the

(M -V + hCG) group and both the (M-V -hCG) and (cytotoxicant, +hCG) groups, which were statistically significant ($p<0.0001$ for both comparisons). The estimated fragment-to-fragment variation in testosterone concentrations was homogeneous among the four control groups (0.2174). The estimated replicate variation in testosterone concentrations was small (0.007214), which is about 2% of total variation.

Based on the analysis of variance results for the log-transformed baseline-adjusted LDH concentrations among the stimulated media-vehicle control group and the three graded AG dose groups, there were differences among the AG groups and the stimulated media-vehicle control (i.e., (M-V +hCG) group), which were borderline significant ($p= 0.0514$). There was a significant negative linear dose trend ($p=0.0235$). As the dose level increased, the log-transformed baseline-adjusted LDH concentrations decreased. The estimated fragment-to-fragment variation in LDH concentrations was homogeneous among the four study groups. The estimated animal variation was small (relative to the replicate variation and fragment-to-fragment variation).

Based on the analysis of variance results for the log-transformed baseline-adjusted LDH concentrations among the four control groups, there was no significant difference among the four control groups ($p=0.4409$). The fragment variation in LDH concentrations varied among the study groups (from 0.002339 in the (M-V -hCG) group to 0.03171 in the (cytotoxicant, +hCG) group). The estimated animal variation was zero.

3.3.5 Lab E

The full Lab E statistical analysis narrative, with graphs and tables, is included in Appendix J.

For the testosterone control group comparisons, two models were run on the data - one assuming heterogeneous fragment to fragment variances among the test groups and one assuming common (homogeneous) fragment to fragment variances among the test groups. Both models yielded similar results in that the overall group differences were significant ($F=75.76$, $p<0.0001$) for the heterogeneous model as well as ($F=74.83$, $p<0.0001$) for the homogeneous assumption. Also, the likelihood ratio test showed no difference between the two models ($p>0.05$). The difference in the log likelihoods was 1.5, which was well under the critical value for this test. A decision was made to choose the simpler homogeneity model to make pairwise group comparisons. M-V-Control w/o hCG and EDS were statistically different from MV-Control+hCG ($p<0.0001$ in both cases). M-Control+hCG was not statistically different from M-V-Control+hCG, $p=0.3269$. The validity of the model was checked by testing the assumption of normality of the residuals with the normal quantile plot. In addition, the Shapiro Lab Ck's test was performed which yielded a $p=0.5716$, thereby indicating that the assumption of normality was not rejected.

For the testosterone MV-Control+hCG and AG comparisons, two models were run on the data - one assuming heterogeneous fragment-to-fragment variances among the test groups and one assuming common (homogeneous) fragment-to-fragment variances among the test groups.

Both models yield similar results in that the overall group differences were significant (F=94.64, p<0.0001) for the heterogeneous model as well as (F=82.12, p<0.0001) for the homogeneous assumption. Also, the likelihood ratio test showed no difference between the two models (p>0.05). The difference in the log likelihoods was 2.1, which was well under the critical value for this test. The simpler homogeneity model was used to make our pairwise group comparisons. All three test groups were statistically different from M-V-Control+hCG (p=0.0055 for AG at 10 µM; p<0.0001 for AG at 100 and 1000 µM). The statistically significant p-values were small enough to accommodate any multiple comparison challenges to these results.

In addition, the linear and quadratic contrasts were evaluated. The linear trend was significant (p<0.0001) but the quadratic contrast was not significant (p=0.2511). Furthermore, any trends of the predicted values of log testosterone were determined over the increasing AG group concentrations. A dramatic decreasing trend was observed from MV-Control+hCG to AG at 1000 µM. Using a simple ANOVA analysis, this was a significant downward trend, R-square=0.993, p=0.0001.

Validity of the model was checked by testing the assumption of normality of the residuals, the normal quantile plot. The Shapiro Lab Ck's test yielded a p=0.4601, which indicated that the assumption of normality was not rejected.

A control chart was performed to determine if the logarithmic process was in control. The logarithmic mean of each testis was plotted over the seven groups to determine if they were within the 3 standard error limits as determined by the variance component model. The process at Lab E appeared to be in control given these liberal limits. However, if the limits were moved to one standard error, then Testes 12, 13 and 18 were out of control.

For LDH control group comparisons, two models were run on the data - one assuming heterogeneous fragment to fragment variances among the test groups and one assuming common (homogeneous) fragment to fragment variances among the test groups. Both models yielded similar results in that the overall group differences were significant (F=8.19, p=0.0006) for the heterogeneous model as well as (F=4.72, p=0.01) for the homogeneous assumption. Also, the likelihood ratio test showed no difference between the two models (p>0.05). The difference in the log likelihoods was1.9 which was well under the critical value for this test. The simpler homogeneity model was used to make pairwise group comparisons. EDS was statistically different from MV-Control+hCG (p=0.0017). The groups, MV-Control-hCG and M-Control+hCG were not statistically different from MV-Control+hCG, p=0.4299 and p=0.3891, respectively. The statistically significant p-value=0.0017 was small enough to accommodate any multiple comparison challenges to these results. The residual plots vs. the replicates and groups did not exhibit any discrepant patterns. The MV-Control+hCG group appeared to have less variation than the others. However, by the standard variance comparisons of Brown-Forsythe, Levene, or Bartlett there were no differences in variance patterns among the groups, p<0.05, which confirmed the choice of the homogeneity model.

Validity of the model was checked by testing the assumption of normality of the residuals, the normal quantile plot. The Shapiro Lab Ck's test yielded a p=0.4558, which

indicated that the assumption of normality was not rejected.

For the LDH MV-Control+hCG and AG comparisons, two models were run on the data - one assuming heterogeneous fragment to fragment variances among the test groups and one assuming common (homogeneous) fragment to fragment variances among the test groups. Both models yielded similar results in that the overall group effect is not significant ($F=1.64$, $p=0.2072$) for the heterogeneous model as well as ($F=1.18$, $p=0.3387$) for the homogeneous assumption. The likelihood ratio test showed no difference between the two models ($p>0.05$). The difference in the log likelihoods was 2.0, which was well under the critical value for this test. The simpler homogeneity model was used to make pairwise group comparisons. However, note that since the overall group effect was not significant, one would ordinarily not proceed with the pairwise group comparisons. However, they were provided for review. The test groups were not statistically different from MV-Control+hCG ($p=0.4710$ for AG at 10 µM; $p=0.1903$ for AG at 100 µM; and $p=0.0907$ for AG at 1000 µM). There was a slight increasing trend from MV-Control+hCG to AG at 1000 µM. Using a simple analysis of variance (ANOVA) analysis, this was not a significant upward trend, R-square=0.159, $p=0.1318$.

The linear trend was not significant ($p=0.0742$), neither was the quadratic contrast. There were no discrepant patterns in the residuals across replicates or groups.

The validity of the model was checked by testing the assumption of normality of the residuals, the normal quantile plot. The Shapiro Lab Ck's test yielded a $p=0.7072$, which indicated that the assumption of normality was not rejected.

The logarithmic mean was plotted for each testis over the seven groups to determine if each mean was within the 3 standard error limits as determined by the variance component model. The process at Lab E appears to be in control given these liberal limits. However, if the limits are moved to one standard error, then Testis 09 was out of control.

3.4 Interlaboratory Statistical Analysis

The interlaboratory statistical analysis results are provided below. The complete statistical analysis narrative is included in Appendix C. Testosterone results are summarized in Tables 3.4-1 and 3.4-2 and illustrated in Figures 3.4-1 to 3.4-8. LDH results are summarized in Tables 3.4-3 and 3.4-4 and illustrated in Figures 3.4-9 to 3.4-16. Coefficient of variation results are summarized and illustrated in Tables 3.4-5 to 3.4-8 and Figures 3.4-17 and 3.4-18.

It is important to note that all laboratories except Lab B used a natural-logarithmic transformation. Lab B used a base 10 logarithmic transformation. Therefore for the interlaboratory analysis the reported values for each endpoint for Lab B were multiplied by ln(10)=2.3026, in order to transform them to the natural-logarithmic scale, to be in conformance with the other laboratories. It should also be noted that results from Lab B were not transformed in the Lab B intralaboratory analysis report, since they are self consistent.

Table 3.4-1 displays the estimated testostosterone concentration values and their associated standard errors and p-values for each of the eight endpoints as reported by the five participating laboratories. These values are based on the intralaboratory analyses and represent the basic information underlying the comparisons among laboratories. The Q-statistics for homogeneity of laboratory results indicate that there is significant heterogeneity among laboratories results for comparison between the (M -V - hCG) control group and the (M-V +hCG) control group (p=0.039). There is a borderline-significant heterogeneity among laboratories for comparison between the AG 1000 µM group and the (M-V +hCG) control group (p=0.058). In these instances of heterogeneity among laboratories, Lab B and Lab D had relatively less reduction from the (M -V + hCG) control group as compared to the other laboratories.

Table 3.4-2 displays the same mean values as Table 3.4-1 and associated within laboratory 95% confidence intervals about these mean values. It also displays the overall mean values and their associated 95% confidence intervals, incorporating among laboratory variation. These means and confidence intervals are displayed in Figures 3.4-1 through 3.4-8. Each figure includes reference lines corresponding to the overall average and to 0. There were significant differences between each of the three graded dose AG groups and the (M-V + hCG) group across all laboratories as well as within laboratories. Significant linear components of trend occurred among the (M-V + hCG) group and the three graded dose AG groups across all laboratories as well as within each laboratory. As the AG concentration level increased, the testosterone concentration levels decreased. Also, quadratic components of trend among the (M-V + hCG) group and the three graded dose AG groups were significant across all laboratories as well as within Lab A, Lab B, and Lab C. Differences between the (M-V + hCG) group and both the (M-V - hCG) and the (Cytotoxicant EDS +hCG) control groups were significant across all laboratories as well as within each laboratory. There were no significant differences between the (M-V + hCG) group and the (Media + hCG) control group across all laboratories or within any laboratory.

Table 3.4-3 displays the estimated LDH concentration values and the associated standard errors and p-values for each of the eight endpoints as reported by the five laboratories. The Q-

statistics for consistency among laboratory results tests show that there was no significant heterogeneity across the individual laboratory results for any of the eight endpoints. Table 3.4-4 displays the same mean values as Table 3.4-3, and associated within laboratory 95% confidence intervals about these mean values. It also displays the overall mean values and their associated 95% confidence intervals, incorporating among laboratory variation. These mean values and confidence intervals are shown in Figures 3.4-9 through 3.4-16. No significant differences existed between any of the AG groups and the (M-V + hCG) groups, either across all laboratories or within each laboratory. A significant linear component of trend occurred among the (M-V + hCG) control group and the three graded dose AG groups within Lab D ($p=0.024$). The difference between the (M-V + hCG) group and the (M-V - hCG) group was significant across all laboratories combined but not within any individual laboratory. The difference between the (M-V + hCG) control group and the (Cytotoxicant EDS +hCG) control group was not significant across all laboratories combined. It was however significant within Laboratory E. Differences between the (M-V + hCG) group and the (Media +hCG) control group were borderline significant across all laboratories combined as well as within Laboratory B.

Appendix C Tables A-1 and A-2 display the within laboratory variance components and associated degrees of freedom for each laboratory. These are the squares of the within laboratory standard errors displayed in Tables 3.4-1 and 3.4-3. Appendix C Tables A-1 and A-2 also display the random laboratory to laboratory variation with associated degrees of freedom and the square of the standard error of the overall mean value. For the two comparisons where Table 3.4-1 indicates significant (or near significant) heterogeneity among the laboratory results, Appendix C Table A-1 shows that the variability among the five laboratories was considerably larger than the variabilities within laboratories. For the LDH results, Appendix C Table A-2 shows that four of the eight endpoints had zero estimated variability among the five laboratories. The other four endpoints had variances among laboratories that were about the same as the variances within laboratories.

Tables 3.4-5 and 3.4-6 display the estimates of the within laboratory coefficients of variation (CV) and their associated within laboratory 95 percent confidence intervals for each of the comparisons for testosterone and LDH concentrations, respectively. They also display the coefficients of variation across laboratories and their associated 95 percent confidence intervals. These overall coefficients of variation and associated 95 percent confidence intervals are shown in Figures 3.4-17 and 3.4-18, for testosterone concentrations and for LDH concentrations respectively. The coefficients of variation among laboratories range from 13% to 44% for testosterone concentrations and from 2.7% to 10.4% for LDH concentrations (depending on effect). The two largest coefficients of variation among laboratories for testosterone concentrations (44% and 39%) occurred for comparisons between the (M - V + hCG) control group and the (M - V - hCG) control group (44%) and between the (M - V + hCG) control group and the (AG 1000 µM) group (39%). For the other comparisons the coefficients of variation among laboratories for testosterone concentrations were less than 30% (Tables 3.4-5 and 3.4-6).

Appendix C Figures A-1 to A-16 display normal probability plots of the logarithmic estimates for the five laboratories for each of the eight endpoints. Appendix C Figures A-1 to A-

8 correspond to testosterone concentrations and Appendix C Figures A-9 to A-16 correspond to LDH concentrations. If the points lie on a smooth curve this indicates that the laboratories are separated randomly (e.g. Appendix C Figures A-6 and A-7). If a subset of the points follow a smooth curve and one or more laboratories are separated from this curve this would suggest that there are systematic differences among some of the laboratories.

Coefficients of variation across laboratories for testosterone concentrations were determined and the results are summarized in Tables 3.4-7 and 3.4-8. The complete statistical analysis report is included in Appendix D. The CVs for individual groups (laboratories) are larger than the CVs for comparisons among groups, which is attributed to using the blocked within testis design for the study. Finally, the CVs based on the logarithmic analysis and those based on the weighted and unweighted exponential analyses are all similar.

Table 3.4-1. Analysis Results Reported by Each Laboratory for Natural Log Transformed Baseline Adjusted Testosterone Concentrations and Tests of Homogeneity of Laboratory Results

Tests	Effect (and Standard Error)[1]					Test for Consistency Results[3]
	Lab A	Lab B[2]	Lab C	Lab D	Lab E	
Comparisons of three AG groups to (MV +hCG) Control Group						
AG 10 µM vs M-V+hCG	-0.689 (0.180) (p=0.001)	-0.407 (0.151) (p=0.012)	-0.414 (0.139) (p=0.007)	-1.041 (0.228) (p=0.001)	-0.699 (0.229) (p=0.005)	Q=2.947 (p=0.567)
AG 100 µM vs M-V+hCG	-2.265 (0.180) (p<0.001)	-1.565 (0.151) (p<0.001)	-1.787 (0.139) (p<0.001)	-2.128 (0.293) (p<0.001)	-2.183 (0.229) (p<0.001)	Q=4.240 (p=0.375)
AG 1000 µM vs M-V+hCG	-3.681 (0.180) (p<0.001)	-2.862 (0.151) (p<0.001)	-3.851 (0.139) (p<0.001)	-3.107 (0.114) (p<0.001)	-3.263 (0.229) (p<0.001)	Q=9.111 (p=0.058)
Dose trends among the AG and (MV +hCG) Groups[4]						
Linear Trend	-1.262 (0.057) (p<0.001)	-0.974 (0.048) (p<0.001)	-1.293 (0.044) (p<0.001)	-1.041 (0.049) (p<0.001)	-1.127 (0.072) (p<0.001)	Q=3.151 (p=0.533)
Quadratic Trend	-0.182 (0.064) (p=0.009)	-0.222 (0.053) (p<0.001)	-0.413 (0.049) (p<0.001)	0.015 (0.091) (p=0.868)	-0.095 (0.081) (p=0.251)	Q=3.273 (p=0.513)
Comparisons of Four Control Groups						
M-V-hCG vs M-V+hCG	-1.985 (0.167) (p<0.001)	-1.420 (0.182) (p<0.001)	-2.408 (0.150) (p<0.001)	-1.289 (0.229) (p<0.001)	-2.137 (0.180) (p<0.001)	Q=10.066 (p=0.039)
Media+hCG vs M-V+hCG	-0.093 (0.167) (p=0.582)	0.093 (0.182) (p=0.614)	0.037 (0.150) (p=0.807)	-0.236 (0.229) (p=0.312)	-0.180 (0.180) (p=0.327)	Q=0.830 (p=0.934)
Cytotoxicant+hCG vs M-V+hCG	-1.491 (0.167) (p<0.001)	-1.307 (0.182) (p<0.001)	-1.360 (0.150) (p<0.001)	-1.398 (0.229) (p<0.001)	-1.816 (0.180) (p<0.001)	Q=1.834 (p=0.766)

1. These results were reported by the five participant laboratories.
2. Lab B used a base 10 logarithmic transformation while the other laboratories used a natural-logarithmic transformation. Therefore the reported values for Lab B were each multiplied by ln(10)=2.3026 to convert them to natural logarithms.
3. Q is defined as $Q=\Sigma(Y_i - Y_w)^2/S_i^2$, where $Y_w = \Sigma[(1/S_i^2) Y_i/(\Sigma 1/S_i^2)]$, and Y_i and S_i are the effect and standard error reported by each participant lab. Under null hypothesis of homogeneous mean effects across labs, Q is approximately a c^2 with 4 (number of labs -1) degree of freedom (DerSimonian and Laird, 1986).
4. Different contrasts were specified among the five labs. In order to use a same contrast for all labs for a given trend test, the following adjustments were made to the reported values (effects and standard errors):
   - for linear dose trends, the reported results by Lab D and Lab B were divided by 10 and 5 respectively;
   - for quadratic trends, the reported results by Lab D and Lab B were divided by 4.

Table 3.4-2. Effects and the 95% Confidence Intervals for Natural Log Transformed Baseline Adjusted Testosterone Concentrations

Test	Effects and 95%Ci[1]					
	Lab A	Lab B[2]	Lab C	Lab D	Lab E	Overall Effect[3,4]
Comparisons of three AG groups to (MV +hCG) Control Group						
AG 10 µM vs M-V+hCG	-0.689 (-1.059 , -0.318)	-0.407 (-0.718 , -0.097)	-0.414 (-0.701 , -0.127)	-1.041 (-1.547 , -0.534)	-0.699 (-1.172 , -0.227)	-0.592 (-0.839 , -0.345)
AG 100 µM vs M-V+hCG	-2.265 (-2.635 , -1.894)	-1.565 (-1.876 , -1.254)	-1.787 (-2.074 ,-1.500)	-2.128 (-2.821 , -1.434)	-2.183 (-2.655 , -1.710)	-1.944 (-2.257 , -1.631)
AG 1000 µM vs M-V+hCG	-3.681 (-4.052 , -3.310)	-2.862 (-3.172 , -2.551)	-3.851 (-4.138 ,-3.564)	-3.107 (-3.366 , -2.848)	-3.263 (-3.735 , -2.790)	-3.350 (-3.774 , -2.926)
Dose Trends among the AG and (MV +hCG) Groups[5]						
Linear Trend	-1.262 (-1.379 , -1.145)	-0.974 (-1.073 , -0.876)	-1.293 (-1.384 , -1.202)	-1.041 (-1.142 , -0.939)	-1.127 (-1.277 , -0.978)	-1.139 (-1.283 , -0.996)
Quadratic Trend	-0.182 (-0.313 , -0.051)	-0.222 (-0.332 , -0.112)	-0.413 (-0.515 , -0.311)	0.015 (-0.178 , 0.208)	-0.095 (-0.262 , 0.072)	-0.193 (-0.355 , -0.031)
Comparisons of Four Control Groups						
M-V-hCG vs M-V+hCG	-1.985 (-2.329 , -1.640)	-1.420 (-1.796 , -1.045)	-2.408 (-2.717 , -2.099)	-1.289 (-1.760 , -0.819)	-2.137 (-2.507 , -1.766)	-1.867 (-2.354 , -1.380)
Media+hCG vs M-V+hCG	-0.093 (-0.438 , 0.251)	0.093 (-0.283 , 0.469)	0.037 (-0.272 , 0.346)	-0.236 (-0.706 , 0.234)	-0.180 (-0.551 , 0.191)	-0.056 (-0.212 , 0.100)
Cytotoxicant+hCG vs M-V+hCG	-1.491 (-1.835 , -1.146)	-1.307 (-1.682 , -0.931)	-1.360 (-1.669, -1.051)	-1.398 (-1.868 , -0.928)	-1.816 (-2.186 , -1.445)	-1.472 (-1.673 , -1.270)

1. The effects and 95% CI were as reported by the five participant laboratories.
2. Lab B used a base 10 logarithmic transformation while the other laboratories used a natural-logarithmic transformation. Therefore the reported values for Lab B were each multiplied by ln(10)=2.3026 to convert them to natural logarithms.
3. The overall effects and standard errors were estimated using a one-way ANOVA mixed model assuming the variances differed among the five labs, where the variances for each lab were fixed to be the reported variances.
4. Degrees of freedom for the (mean) overall effect variance were estimated by $2*((1/K)*\text{å}\ (S_L^2 + S_i^2))^2/(\text{var}(S_L^2) + (2/K^2)*\ \text{å}\ (S_i^4/df_i))$, where S_i^2 is random lab variance, S_i^2 and df_i are reported variance and degree of freedom for a given laboratory, $\text{var}(S_L^2)$ is the variance associated with the estimation of S_L^2, and K is the number of laboratories (Hartung and Makambi, 2001).
5. Different contrasts were specified among the five labs. In order to use a same contrast for all labs for a given trend test, the following adjustments were made to the reported values (effects and standard errors):
   - for linear dose trends, the reported results by Lab D and Lab B were divided by 10 and 5 respectively;
   - for quadratic trends, the reported results by Lab D and Lab B were divided by 4.

Table 3.4-3. Analysis Results Reported by Each Laboratory for Natural Log Transformed Baseline Adjusted LDH Concentrations and Tests of Homogeneity of Laboratory Results

Tests	Effect (and Standard Error)[1]					Test for Consistency Results[3]
	Lab A	Lab B[2]	Lab C	Lab D	Lab E	
Comparisons of three AG groups to (MV +hCG) Control Group						
AG 10 µM vs M-V+hCG	-0.114 (0.128) (p=0.383)	-0.259 (0.200) (p=0.207)	0.058 (0.064) (p=0.373)	0.029 (0.063) (p=0.652)	0.049 (0.067) (p=0.471)	Q=1.152 (p=0.886)
AG 100 µM vs M-V+hCG	0.142 (0.128) (p=0.280)	-0.076 (0.167) (p=0.652)	-0.010 (0.064) (p=0.877)	-0.115 (0.063) (p=0.082)	0.090 (0.067) (p=0.190)	Q=1.048 (p=0.902)
AG 1000 µM vs M-V+hCG	-0.006 (0.128) (p=0.965)	-0.148 (0.167) (p=0.384)	0.026 (0.064) (p=0.688)	-0.112 (0.063 (p=0.088)	0.117 (0.067) (p=0.091)	Q=1.093 (p=0.895)
Dose Trends among the AG and (MV +hCG) Groups[4]						
Linear Trend	0.024 (0.039) (p=0.550)	-0.026 (0.052) (p=0.616)	0.001 (0.020) (p=0.961)	-0.048(0.020) (p=0.024)	0.039 (0.021) (p=0.074)	Q=0.432 (p=0.980)
Quadratic Trend	-0.008 (0.045) (p=0.852)	0.047 (0.051) (p=0.369)	-0.005 (0.023) (p=0.827)	-0.007 (0.022) (p=0.768)	-0.005 (0.024) (p=0.824)	Q=0.100 (p=0.999)
Comparisons of Four Control Groups						
M-V-hCG vs M-V+hCG	-0.038 (0.104) (p=0.721)	-0.057 (0.245) (p=0.819)	0.091 (0.051) (p=0.088)	0.056 (0.041) (p=0.196)	0.050 (0.062) (p=0.430)	Q=0.324 (p=0.988)
Media+hCG vs M-V+hCG	-0.011 (0.104) (p=0.915)	0.190 (0.093) (p=0.052)	0.062 (0.051) (p=0.238)	0.025 (0.057) (p=0.675)	0.054 (0.062) (p=0.389)	Q=0.509 (p=0.973)
Cytotoxicant+hCG vs M-V+hCG	0.186 (0.104) (p=0.085)	-0.080 (0.125) (p=0.526)	0.091 (0.051) (p=0.088)	-0.015 (0.070) (p=0.835)	0.219 (0.062) (p=0.002)	Q=1.499 (p=0.827)

1. These results were reported by the five participant laboratories.
2. Lab B used a base 10 logarithmic transformation while the other laboratories used a natural-logarithmic transformation. Therefore the reported values for Lab B were each multiplied by ln(10)=2.3026 to convert them to natural logarithms.
3. Q is defined as $Q=\sum(Y_i - Y_w)^2/S_i^2$, where $Y_w=\Sigma[(1/S_i^2)\,Y_i/(\Sigma 1/S_i^2)]$, and Y_i and S_i are the effect and standard error reported by each participant lab . Under null hypothesis of homogeneous mean effects across labs, Q is approximately a c^2 with 4 (number of labs -1) degree of freedom (DerSimonian and Laird, 1986).
4. Different contrasts were specified among the five labs. In order to use a same contrast for all labs for a given trend test, the following adjustments were made to the reported values (effects and standard errors):
   - for linear dose trends, the reported results by Lab D and Lab B were divided by 10 and 5 respectively;
   - for quadratic trends, the reported results by Lab D and Lab B were divided by 4.

Table 3.4-4. Effects and the 95% Confidence Intervals for Natural Log Transformed Baseline Adjusted LDH Concentrations

Tests	Effects and 95%CI[1]					
	Lab A	Lab B[2]	Lab C	Lab D	Lab E	Overall Effect[3,4]
Comparisons of Three AG Groups to (MV +hCG) Control Group						
AG 10 µM vs M-V+hCG	-0.114 (-0.378 , 0.151)	-0.259 (-0.671 , 0.153)	0.058 (-0.073 , 0.189)	0.029 (-0.101 , 0.159)	0.049 (-0.089 , 0.186)	0.023 (-0.047 , 0.094)
AG 100 µM vs M-V+hCG	0.142 (-0.123 , 0.406)	-0.076 (-0.421 , 0.268)	-0.010 (-0.141 , 0.121)	-0.115 (-0.245 , 0.015)	0.090 (-0.048 , 0.227)	-0.002 (-0.096 , 0.092)
AG 1000 µM vs M-V+hCG	-0.006 (-0.270 , 0.259)	-0.148 (-0.494 , 0.197)	0.026 (-0.105 , 0.157)	-0.112 (-0.243 , 0.018)	0.117 (-0.020 , 0.254)	-0.004 (-0.104 , 0.097)
Dose Trends among the AG and (MV +hCG) Groups[5]						
Linear Trend	0.024 (-0.057 , 0.105)	-0.026 (-0.133 , 0.080)	0.001 (-0.040 , 0.042)	-0.048 (-0.089 , -0.007)	0.039 (-0.004 , 0.083)	-0.002 (-0.039 , 0.035)
Quadratic Trend	-0.008 (-0.100 , 0.084)	0.047 (-0.059 , 0.152)	-0.005 (-0.051 , 0.041)	-0.007 (-0.052 , 0.039)	-0.005 (-0.054 , 0.043)	-0.003 (-0.027 , 0.021)
Comparisons of Four Control Groups						
M-V-hCG vs M-V+hCG	-0.038 (-0.252 , 0.177)	-0.057 (-0.562 , 0.449)	0.091 (-0.015 , 0.197)	0.056 (-0.034 , 0.146)	0.050 (-0.078 , 0.178)	0.057 (0.002 , 0.112)
Media+hCG vs M-V+hCG	-0.011 (-0.225 , 0.203)	0.190 (-0.001 , 0.382)	0.062 (-0.044 , 0.168)	0.025 (-0.097 , 0.146)	0.054 (-0.074 , 0.182)	0.057 (-0.001 , 0.116)
Cytotoxicant+hCG vs M-V+hCG	0.186 (-0.028 , 0.401)	-0.080 (-0.338 , 0.177)	0.091 (-0.015 ,0.197)	-0.015 (-0.164 , 0.135)	0.219 (0.091 , 0.347)	0.095 (-0.014 , 0.203)

1. The effects and 95% CI were as reported by the five participant laboratories.
2. Lab B used a base 10 logarithmic transformation while the other laboratories used a natural-logarithmic transformation. Therefore the reported values for Lab B were each multiplied by ln(10)=2.3026 to convert them to natural logarithms.
3. The overall effects and standard errors were estimated using a one-way ANOVA mixed model assuming the variances differed among the five labs, where the variances for each lab were fixed to be the reported variances.
4. Degrees of freedom for the (mean) overall effect variance were estimated by $2*((1/K)*\sum (S_l^2 + S_i^2))^2/(var(S_l^2) +(2/K^2)* \sum (S_i^4/df_i))$, where S_l^2 is random lab variance, S_i^2 and df_i are reported variance and degree of freedom for a given laboratory, $var(S_L^2)$ is the variance associated with the estimation of S_L^2, and K is the number of laboratories (Hartung and Makambi, 2001).
5. Different contrasts were specified among the five labs. In order to use a same contrast for all labs for a given trend test, the following adjustments were made to the reported values (effects and standard errors):
   - for linear dose trends, the reported results by Lab D and Lab B were divided by 10 and 5 respectively;
   - for quadratic trends, the reported results by Lab D and Lab B were divided by 4.

Table 3.4-5. Coefficient of Variation and 95% Confidence Intervals for Natural Log Transformed Baseline Adjusted Testosterone Concentrations (%)

Test	CV and 95% CI[1]					
	Lab A	Lab B[2]	Lab C	Lab D	Lab E	Overall
Comparisons of Three AG Groups to (MV +hCG) Control						
AG 10 µM vs M-V+hCG	18.12 (14.10 , 25.39)	15.14 (11.80 , 21.18)	13.98 (10.89 , 19.53)	23.10 (16.10 , 41.20)	23.20 (18.02 , 32.68)	21.43 (13.16 , 57.80)
AG 100 µM vs M-V+hCG	18.12 (14.10 , 25.39)	15.14 (11.80 , 21.18)	13.98 (10.89 , 19.53)	29.94(19.54 , 65.51)	23.20 (18.02 , 32.68)	29.48 (18.89 , 68.78)
AG 1000 µM vs M-V+hCG	18.12 (14.10 , 25.39)	15.14 (11.80 , 21.18)	13.98 (10.89 , 19.53)	11.46 (7.85 , 21.22)	23.20 (18.02 , 32.68)	39.10 (24.21 , 108.48)
Dose Trends among the AG and (MV +hCG) Groups[3]						
Linear Trend	5.69 (4.44 , 7.92)	4.76 (3.72 , 6.63)	4.40 (3.44 , 6.13)	4.86 (3.72 , 7.03)	7.25 (5.66 , 10.10)	12.81 (8.09 , 30.52)
Quadratic Trend	6.36 (4.96 , 8.86)	5.33 (4.16 , 7.41)	4.92 (3.84 , 6.85)	9.10 (6.74 , 14.00)	8.11 (6.33 , 11.30)	14.27 (8.93 , 35.27)
Comparisons of Four Control Groups						
M-V-hCG vs M-V+hCG	16.82 (13.10 , 23.55)	18.36 (14.29 , 25.74)	15.04 (11.72 , 21.04)	23.17 (18.15 , 32.15)	18.12 (14.10 , 25.39)	44.15 (26.74 , 140.41)
Media+hCG vs M-V+hCG	16.82 (13.10 , 23.55)	18.36 (14.29 , 25.74)	15.04 (11.72 , 21.04)	23.17(18.15 , 32.15)	18.12 (14.10 , 25.39)	17.77 (15.71 , 20.48)
Cytotoxicant+hCG vs M-V+hCG	16.82 (13.10 , 23.55)	18.36 (14.29 , 25.74)	15.04 (11.72 , 21.04)	23.17 (18.15 , 32.15)	18.12 (14.10 , 25.39)	18.06 (11.36 , 44.00)

1. The coefficient of variation (CV) was defined as the square root of (exponential variance -1), and lower and upper limits of the 95% CI for CV were given by square root of (exponential of $(df*S^2/\chi^{-1}_{df,\ 0.975}) - 1$) and square root of (exponential of $(df*S^2/\chi^{-1}_{df,\ 0.025}) - 1$) respectively, where S^2 and df were given in Table A-1 of Appendix C.
2. Lab B used a base 10 logarithmic transformation while the other laboratories used a natural-logarithmic transformation. Therefore the reported values for Lab B were each multiplied by ln(10)=2.3026 to convert them to natural logarithms.
3. Different contrasts were specified among the five labs. In order to use a same contrast for all labs for a given trend test, the following adjustments were made to the reported values (effects and standard errors):
   - for linear dose trends, the reported results by Lab D and Lab B were divided by 10 and 5 respectively;
   - for quadratic trends, the reported results by Lab D and Lab B were divided by 4.

Table 3.4-6. Coefficient of Variation and 95% Confidence Intervals for Natural Log Transformed Baseline Adjusted LDH Concentrations (%)

Test	CV and 95% CI[1]					
	Lab A	Lab B[2]	Lab C	Lab D	Lab E	Overall
Comparisons of Three AG Groups to (MV +hCG) Control Group						
AG 10 µM vs M-V+hCG	12.87 (10.04 , 17.98)	20.16 (15.68 , 28.32)	6.42 (5.13 , 8.58)	6.35 (5.01 , 8.69)	6.66 (5.20 , 9.27)	7.90 (6.70 , 9.63)
AG 100 µM vs M-V+hCG	12.87 (10.04 , 17.98)	16.82 (13.10 , 23.55)	6.42 (5.13 , 8.58)	6.35 (5.01 , 8.69)	6.66 (5.20 , 9.27)	9.73 (7.02 , 15.85)
AG 1000 µM vs M-V+hCG	12.87 (10.04 , 17.98)	16.85 (13.12 , 23.60)	6.42 (5.13 , 8.58)	6.35 (5.01 , 8.69)	6.66 (5.20 , 9.27)	10.36 (7.46 , 16.97)
Dose Trends among the AG and (MV +hCG) Groups[3]						
Linear Trend	3.94 (3.07 , 5.48)	5.16 (4.03 , 7.18)	2.03 (1.62 , 2.71)	2.00 (1.58 , 2.73)	2.10 (1.64 , 2.93)	3.74 (2.62 , 6.52)
Quadratic Trend	4.46 (3.48 , 6.21)	5.11 (3.99 , 7.12)	2.27 (1.81 , 3.03)	2.20 (1.74 , 3.01)	2.35 (1.84 , 3.27)	2.73 (2.37 , 3.24)
Comparisons of Four Control Groups						
M-V-hCG vs M-V+hCG	10.41 (8.12 , 14.52)	24.87 (19.30 , 35.09)	5.12 (4.00 , 7.13)	4.06 (2.85, 7.04)	6.21 (4.84 , 8.64)	6.07 (4.98 , 7.78)
Media+hCG vs M-V+hCG	10.41 (8.12 , 14.52)	9.30 (7.26 , 12.96)	5.12 (4.00 , 7.13)	5.75 (4.28, 8.78)	6.21 (4.84 , 8.64)	6.59 (5.75 , 7.73)
Cytotoxicant+hCG vs M-V+hCG	10.41 (8.12 , 14.52)	12.52 (9.76 , 17.48)	5.12 (4.00 , 7.13)	7.02 (5.17 , 10.94)	6.21 (4.84 , 8.64)	10.04 (6.51 , 21.83)

1. The coefficient of variation (CV) was defined as the square root of (exponential variance -1), and lower and upper limits of the 95% CI for CV were given by square root of (exponential of ($df*S^2/\chi^{-1}_{df, 0.975}$) - 1) and square root of (exponential of ($df*S^2/\chi^{-1}_{df, 0.025}$) - 1) respectively, where S^2 and df were given in Table A-2 of Appendix C.

2. Lab B used a base 10 logarithmic transformation while the other laboratories used a natural-logarithmic transformation. Therefore the reported values for Lab B were each multiplied by ln(10)=2.3026 to convert them to natural logarithms.

3. Different contrasts were specified among the five labs. In order to use a same contrast for all labs for a given trend test, the following adjustments were made to the reported values (effects and standard errors):

   · for linear dose trends, the reported results by Lab D and Lab B were divided by 10 and 5 respectively;

   · for quadratic trends, the reported results by Lab D and Lab B were divided by 4.

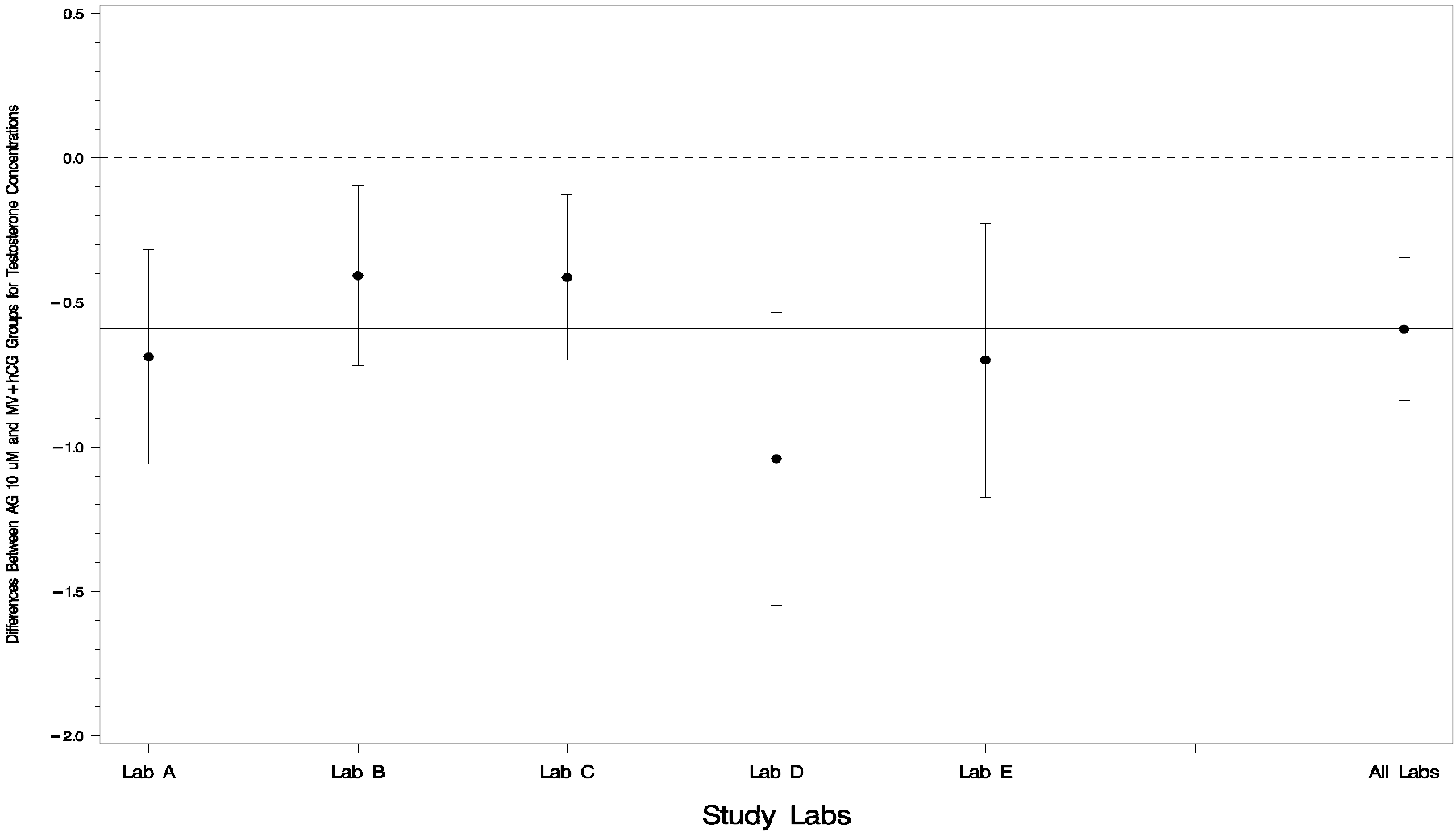

Figure 3.4-1. Differences Between AG 10 µM and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory.

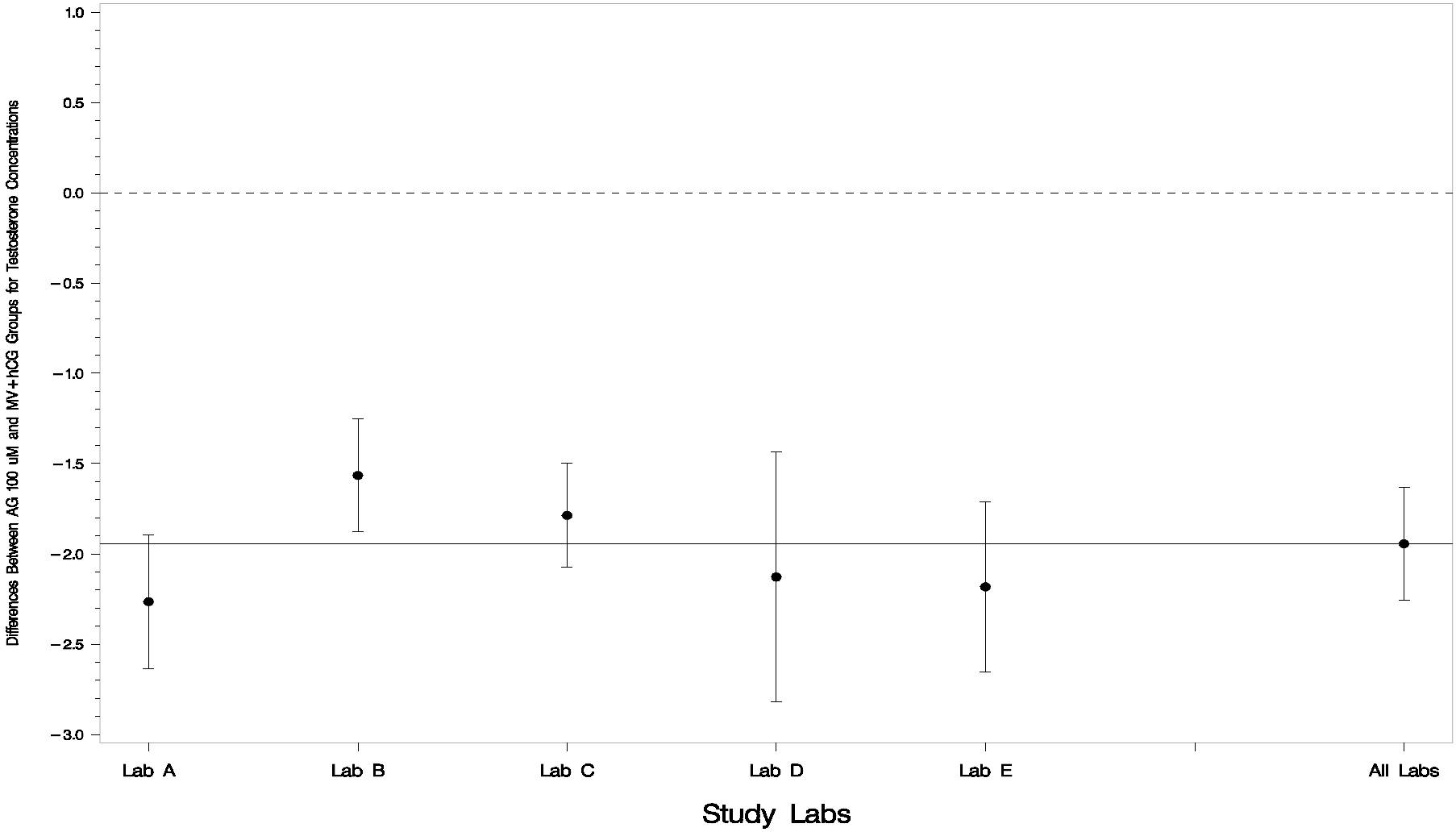

Figure 3.4-2. Differences Between AG 100 µM and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory.

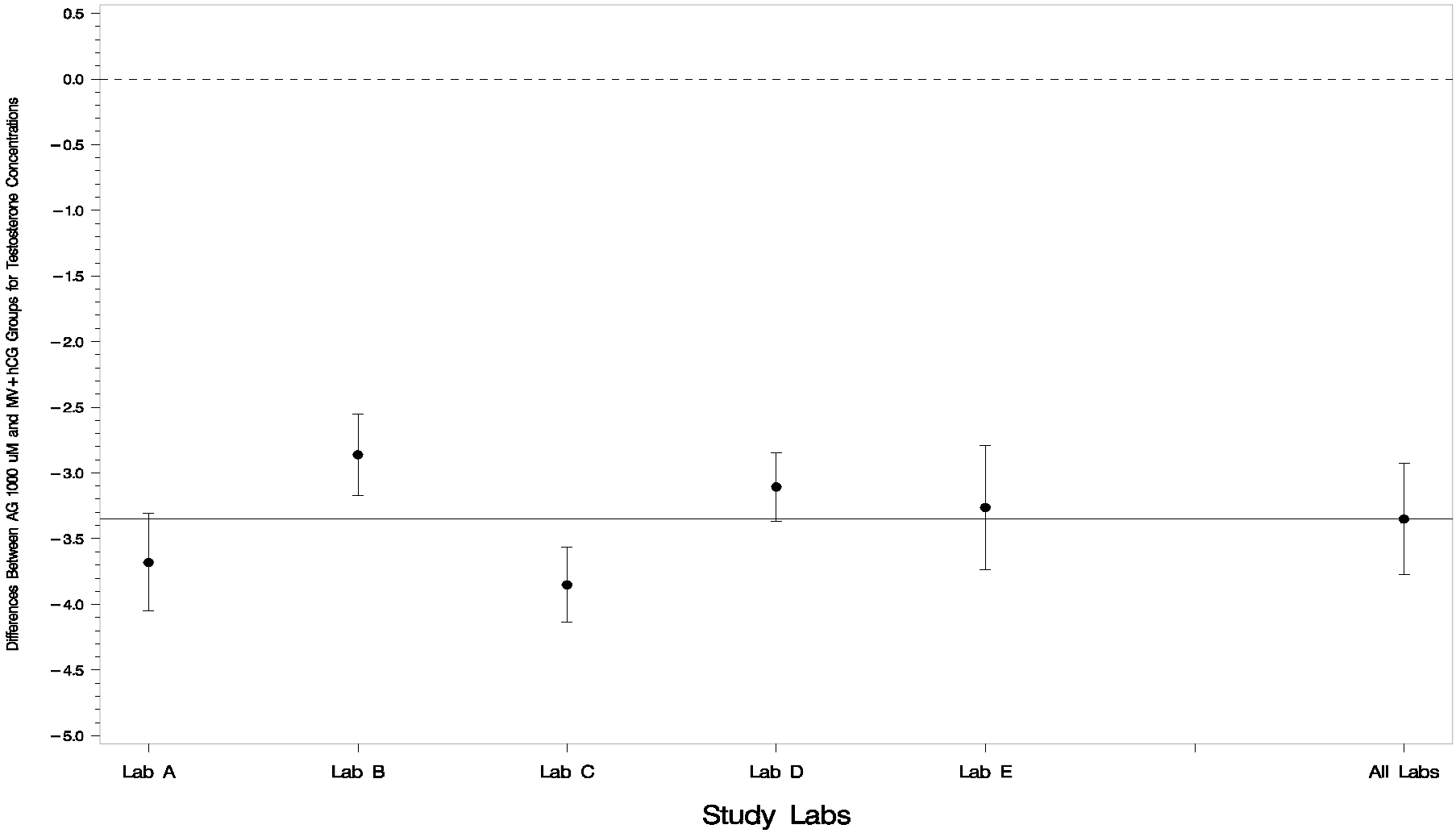

Figure 3.4-3. Differences Between AG 1000 µM and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory

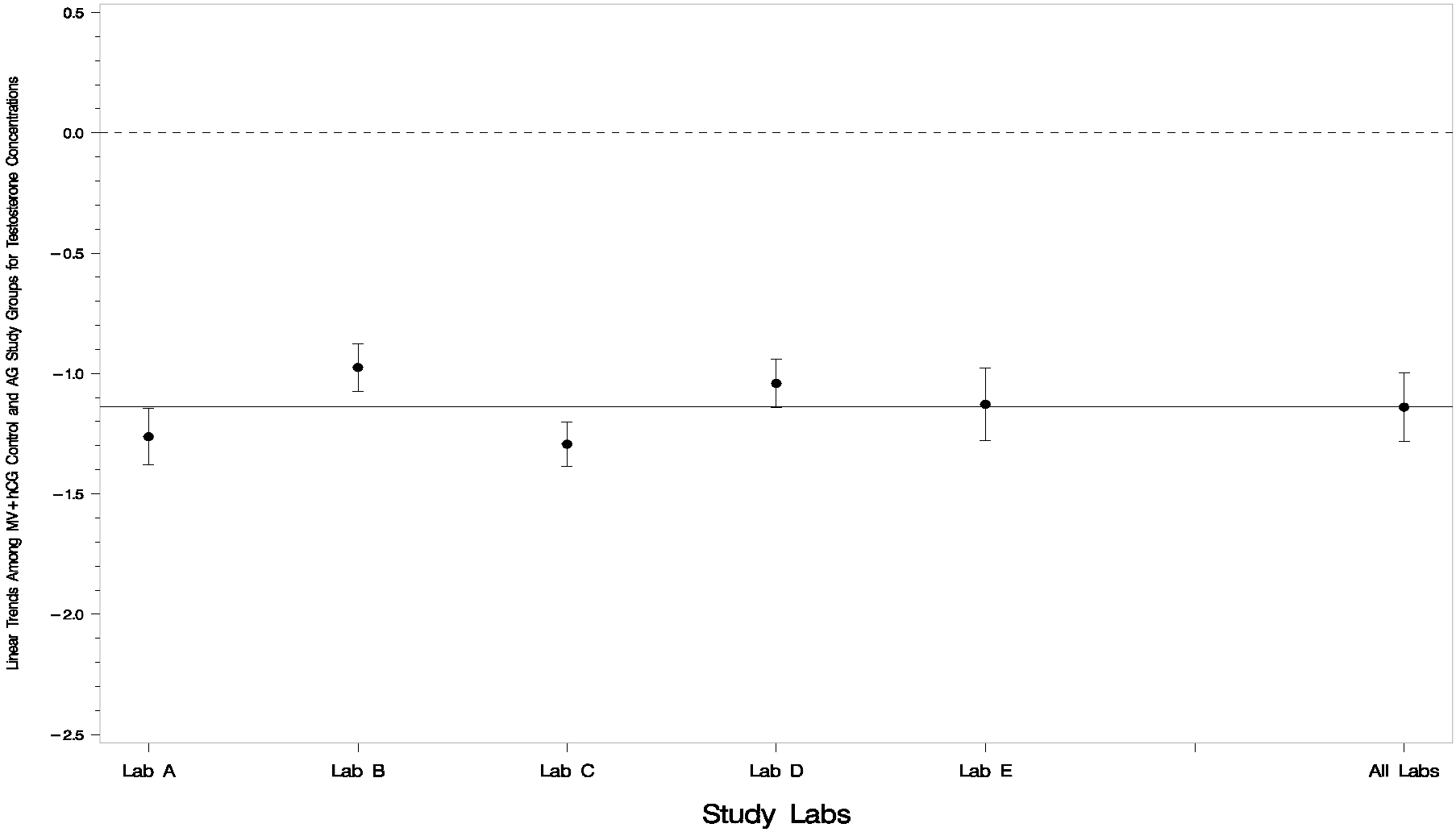

Figure 3.4-4. Linear Trends Among the M-V +hCG Control and the Three AG Study Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory

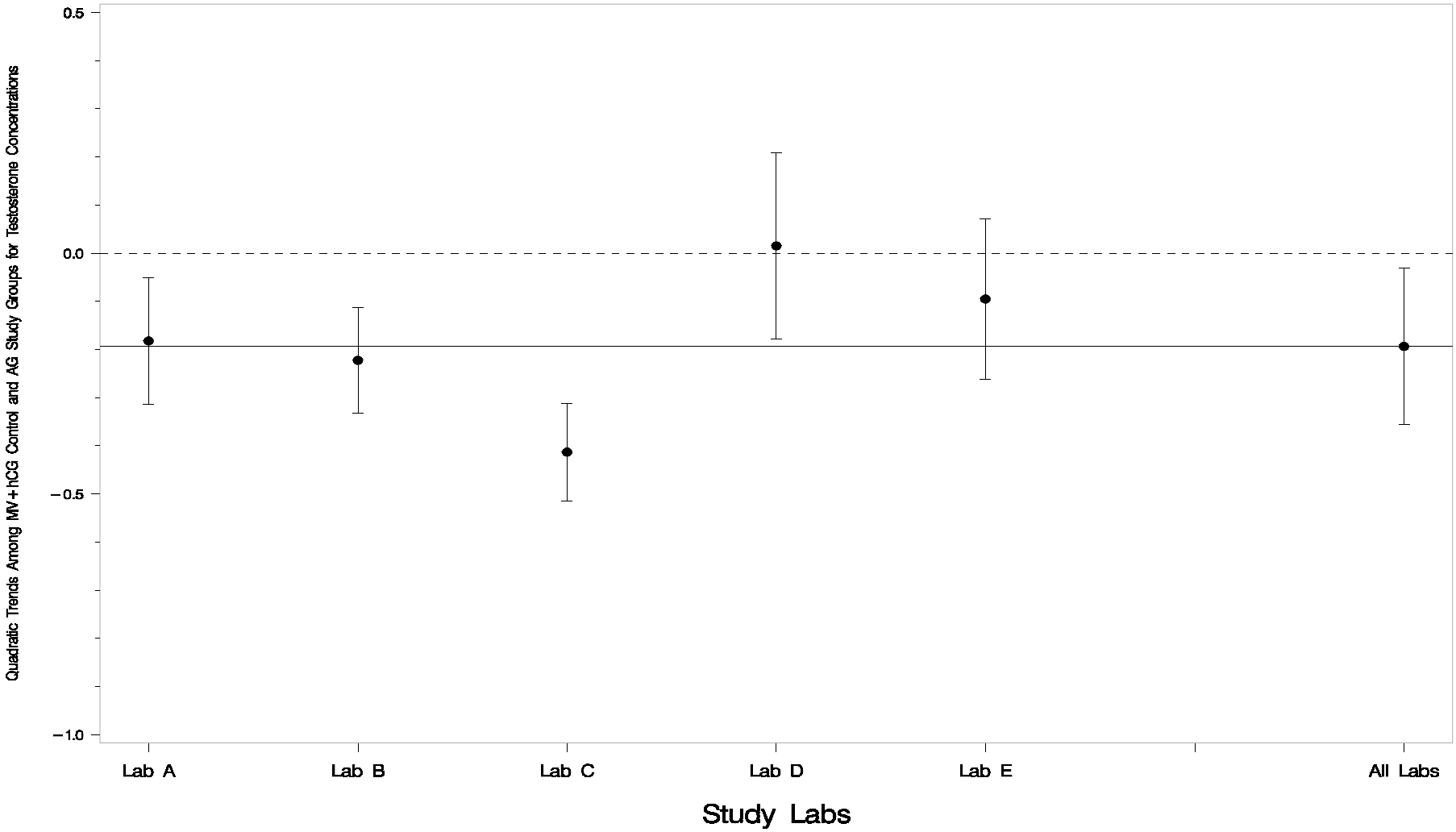

Figure 3.4-5. Quadratic Trends Among the M-V +hCG Control and the Three AG Study Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory

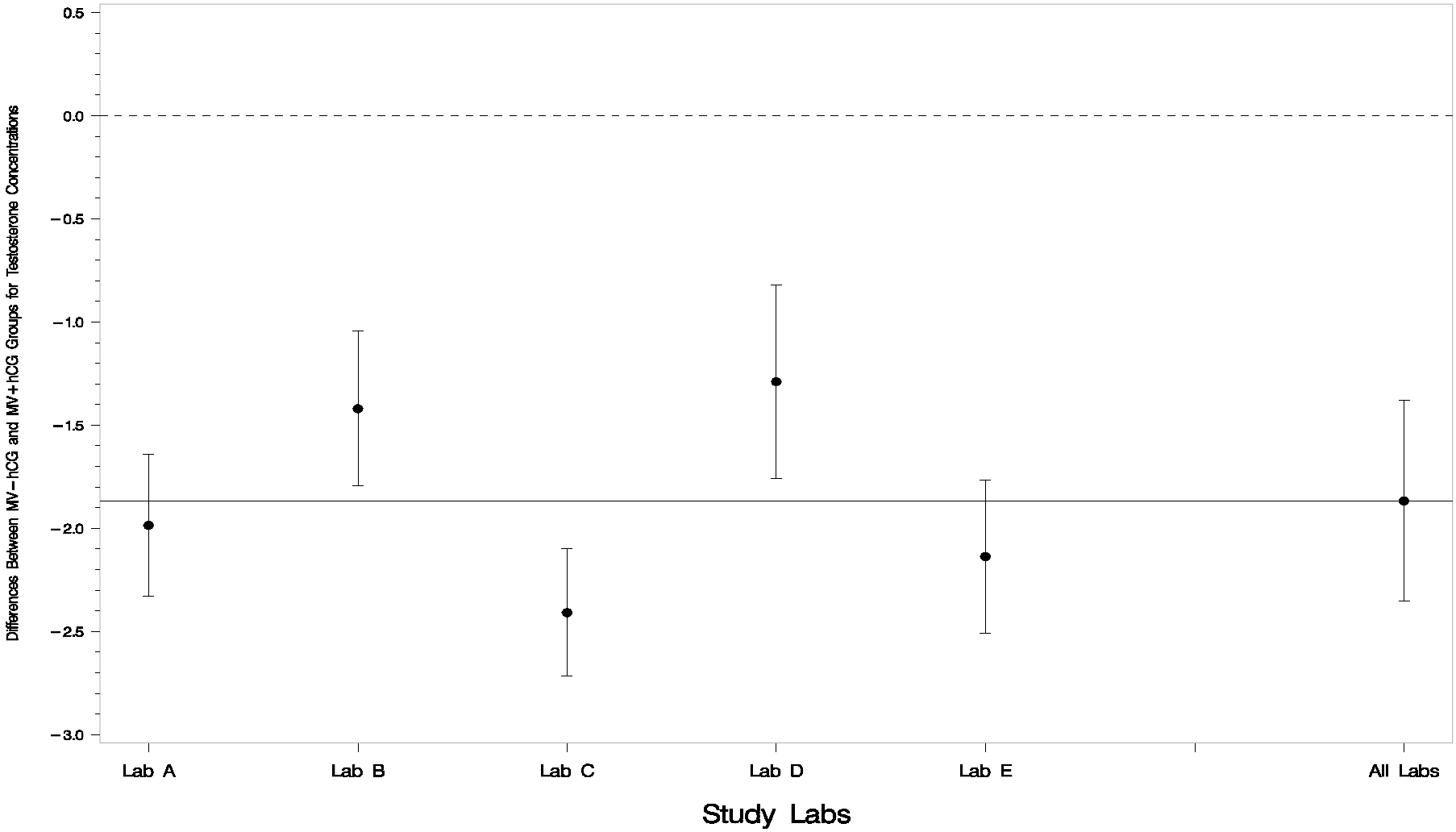

Figure 3.4-6. Differences Between M-V –hCG and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory

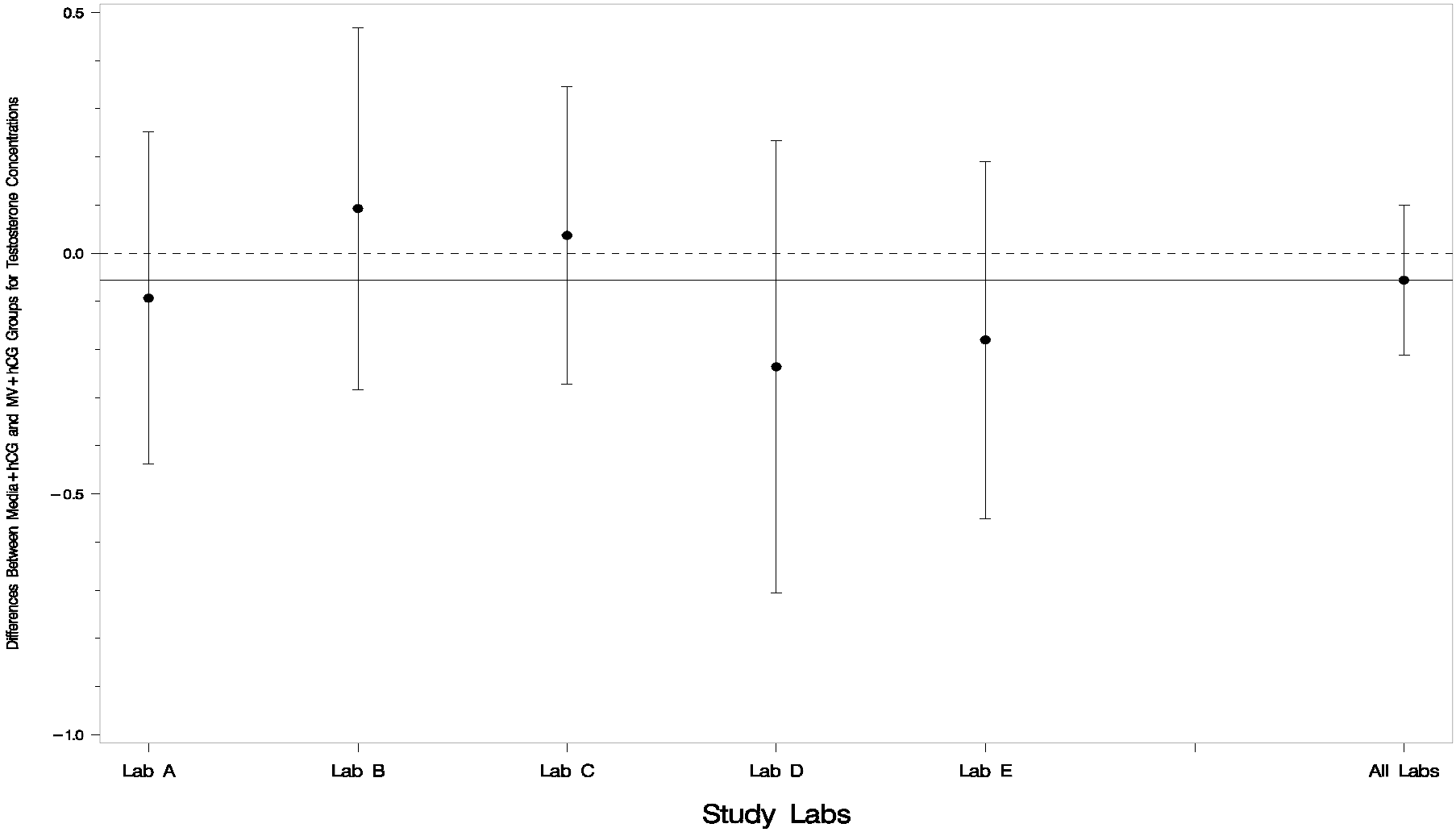

Figure 3.4-7. Differences Between Media+hCG and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory

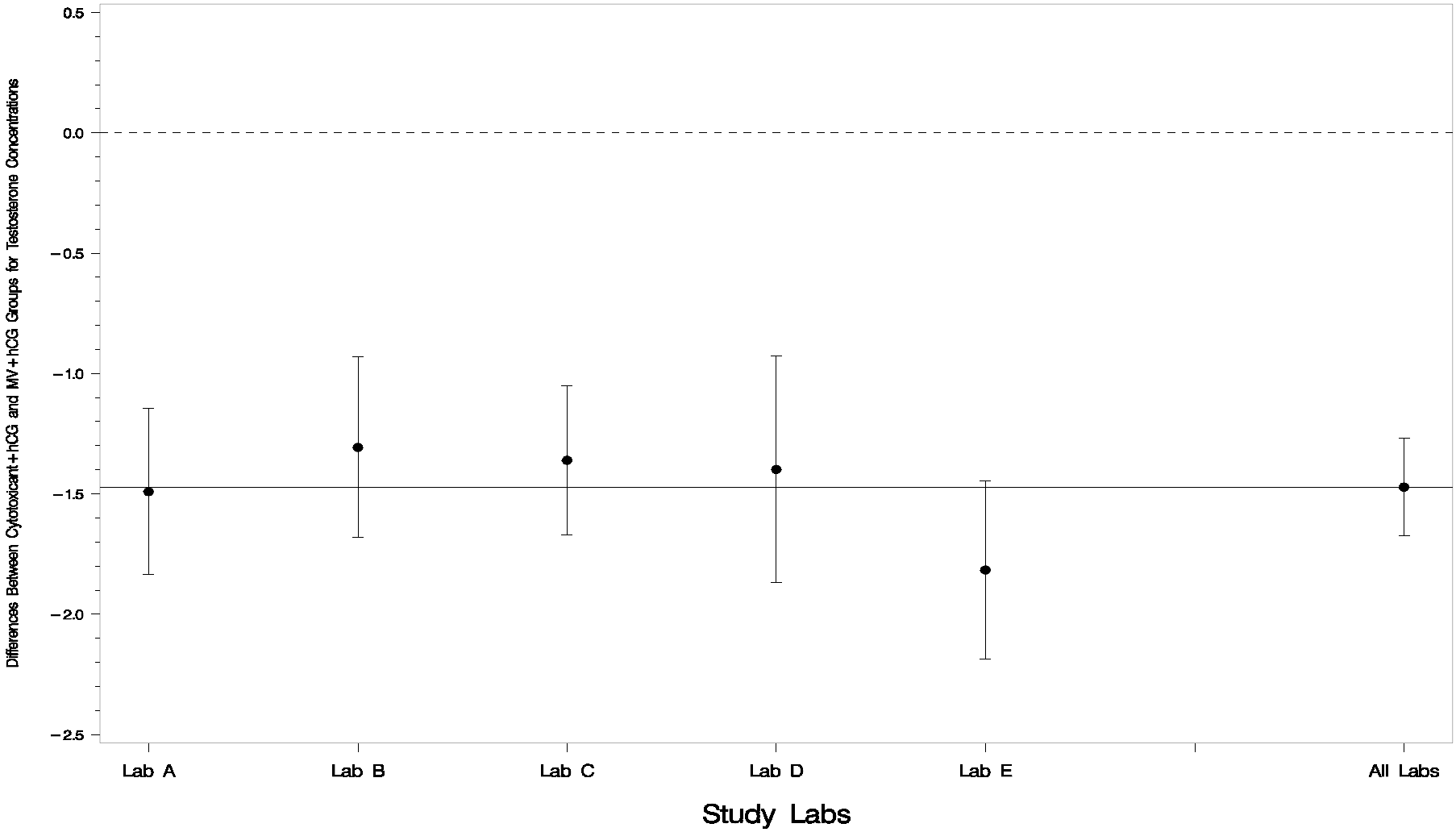

Figure 3.4-8. Differences Between Cytotoxicant+hCG and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations, Across Laboratories and by Each Laboratory

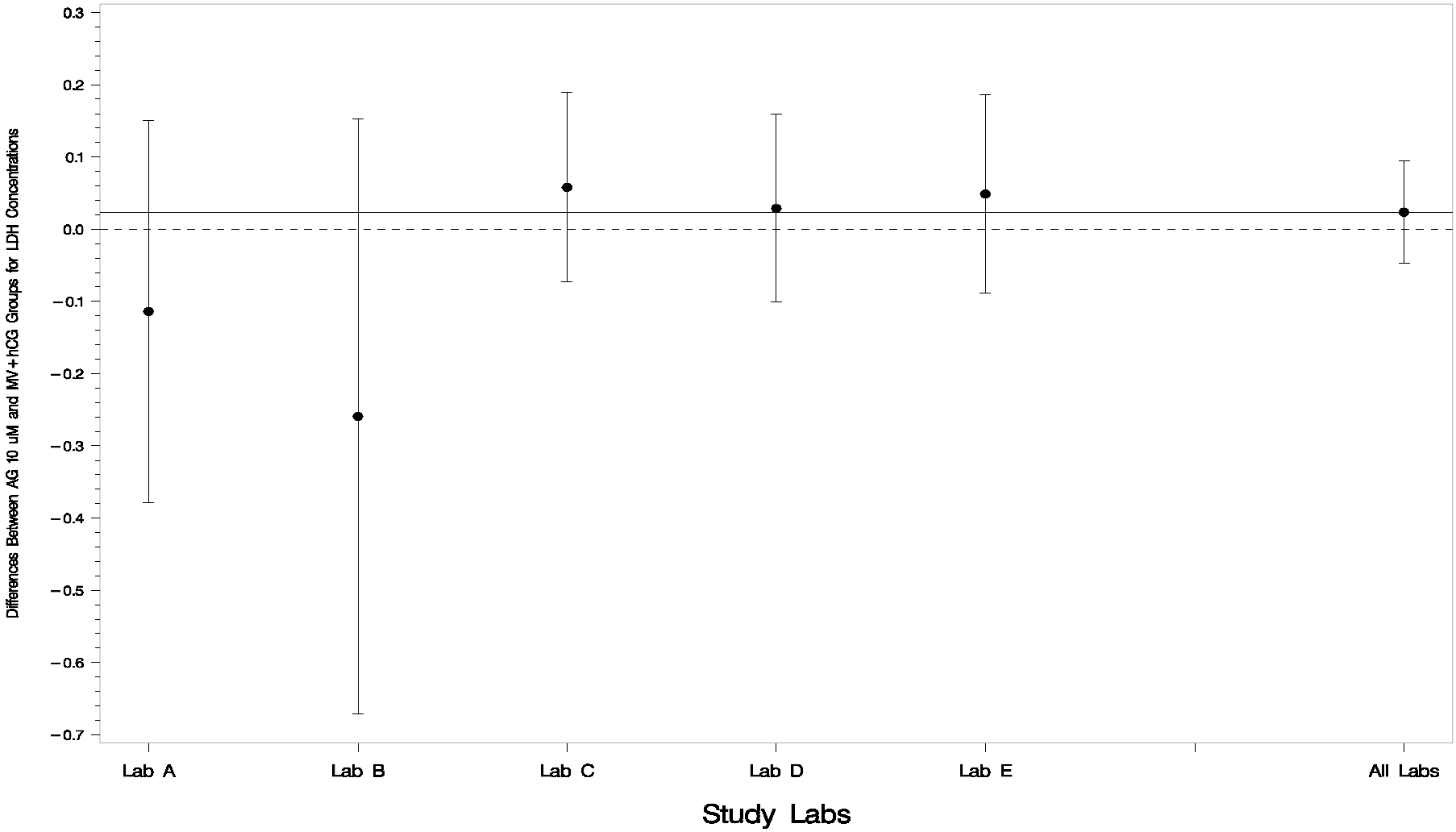

Figure 3.4-9. Differences Between AG 10 µM and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

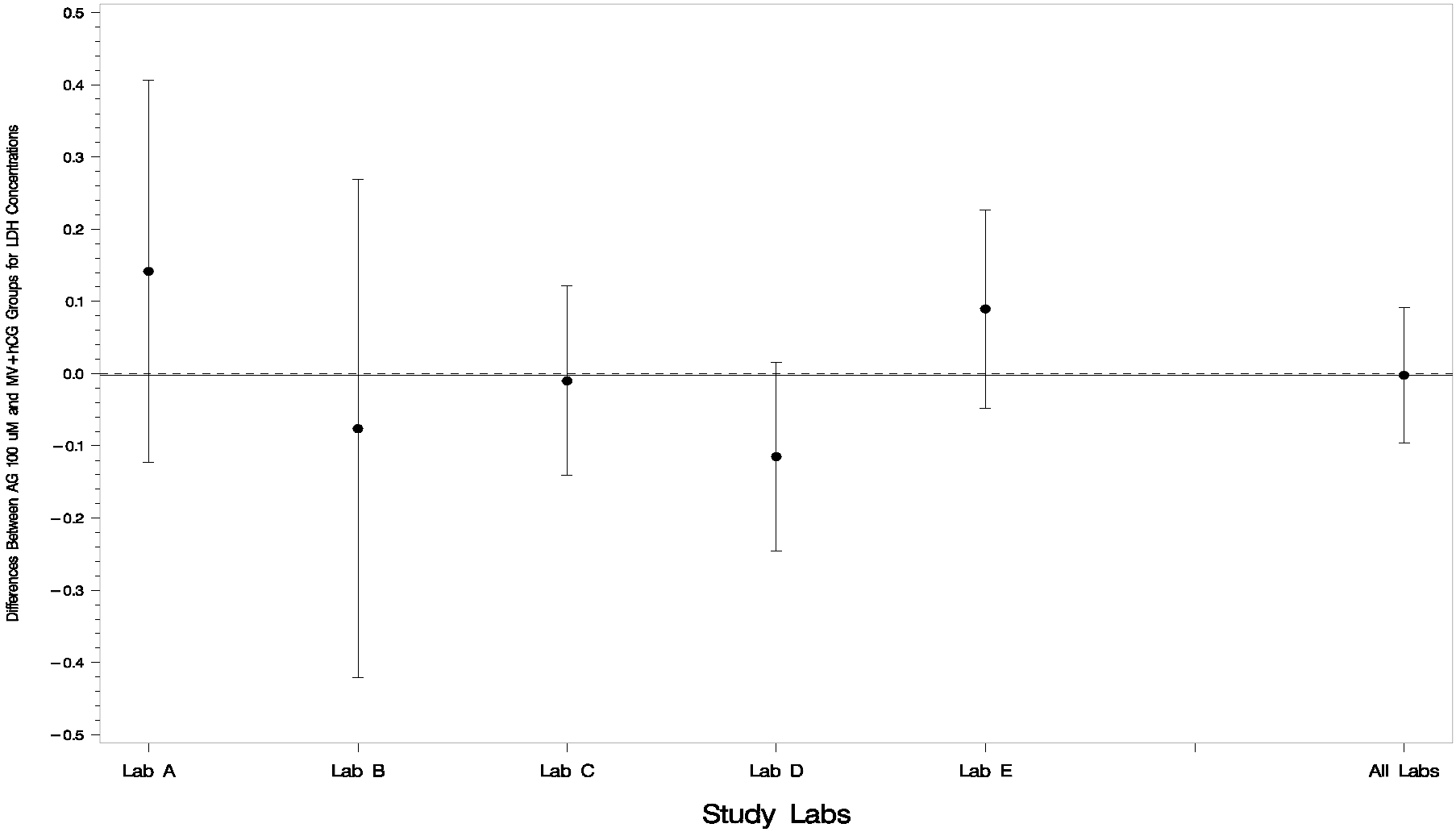

Figure 3.4-10. Differences Between AG 100 µM and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

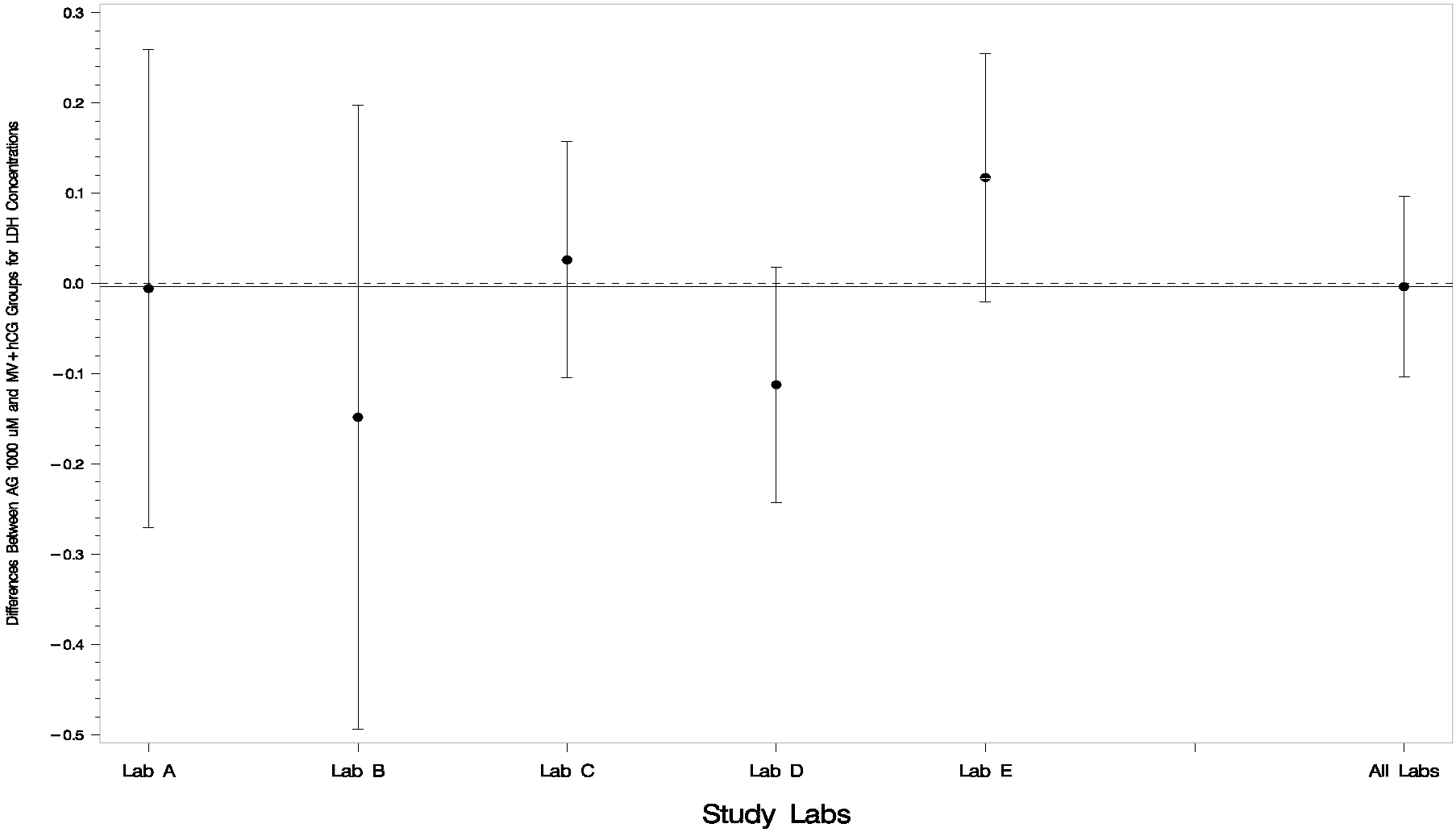

Figure 3.4-11. Differences Between AG 1000 µM and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

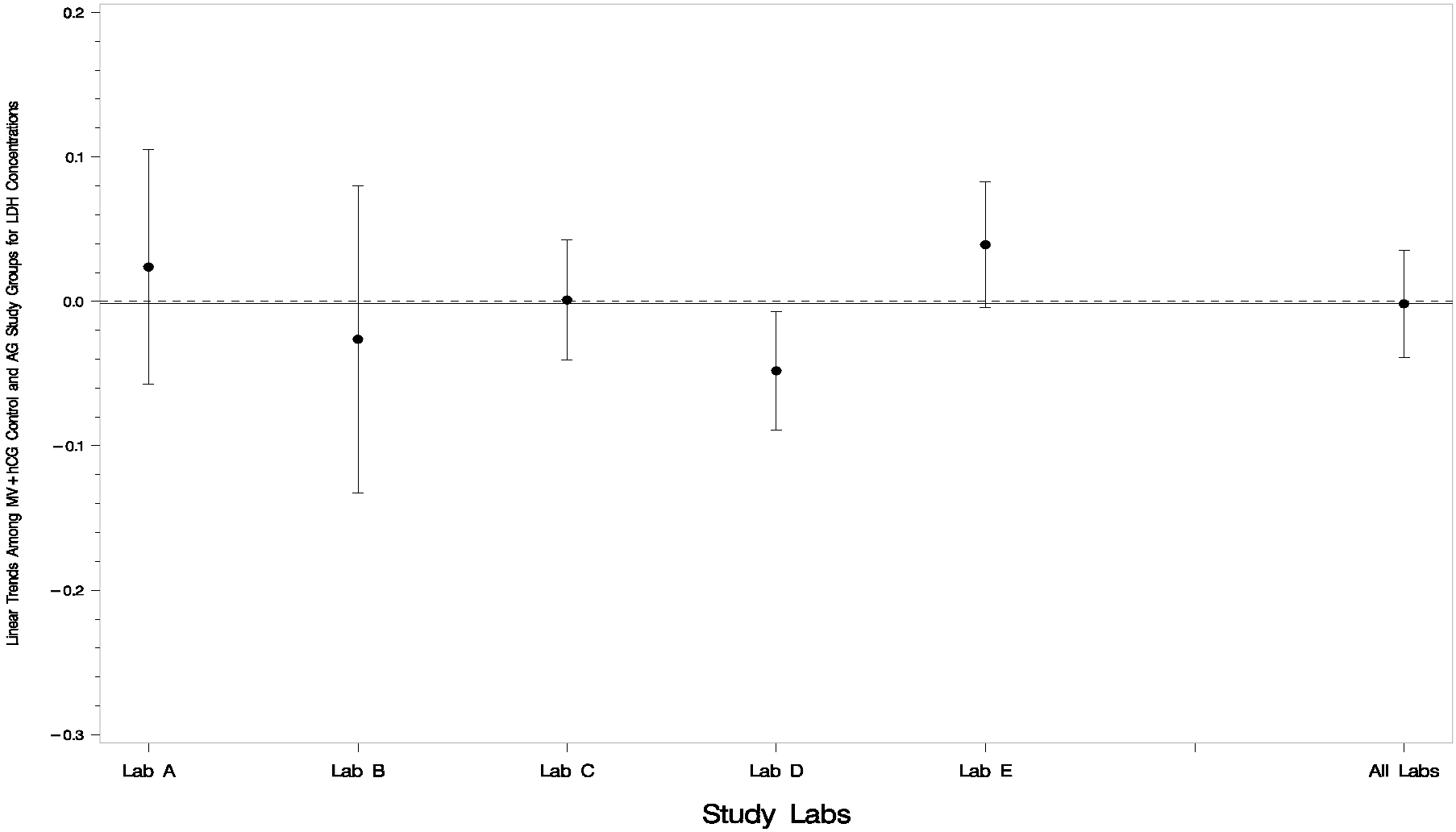

Figure 3.4-12. Linear Trends Among the M-V +hCG Control and the Three AG Study Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

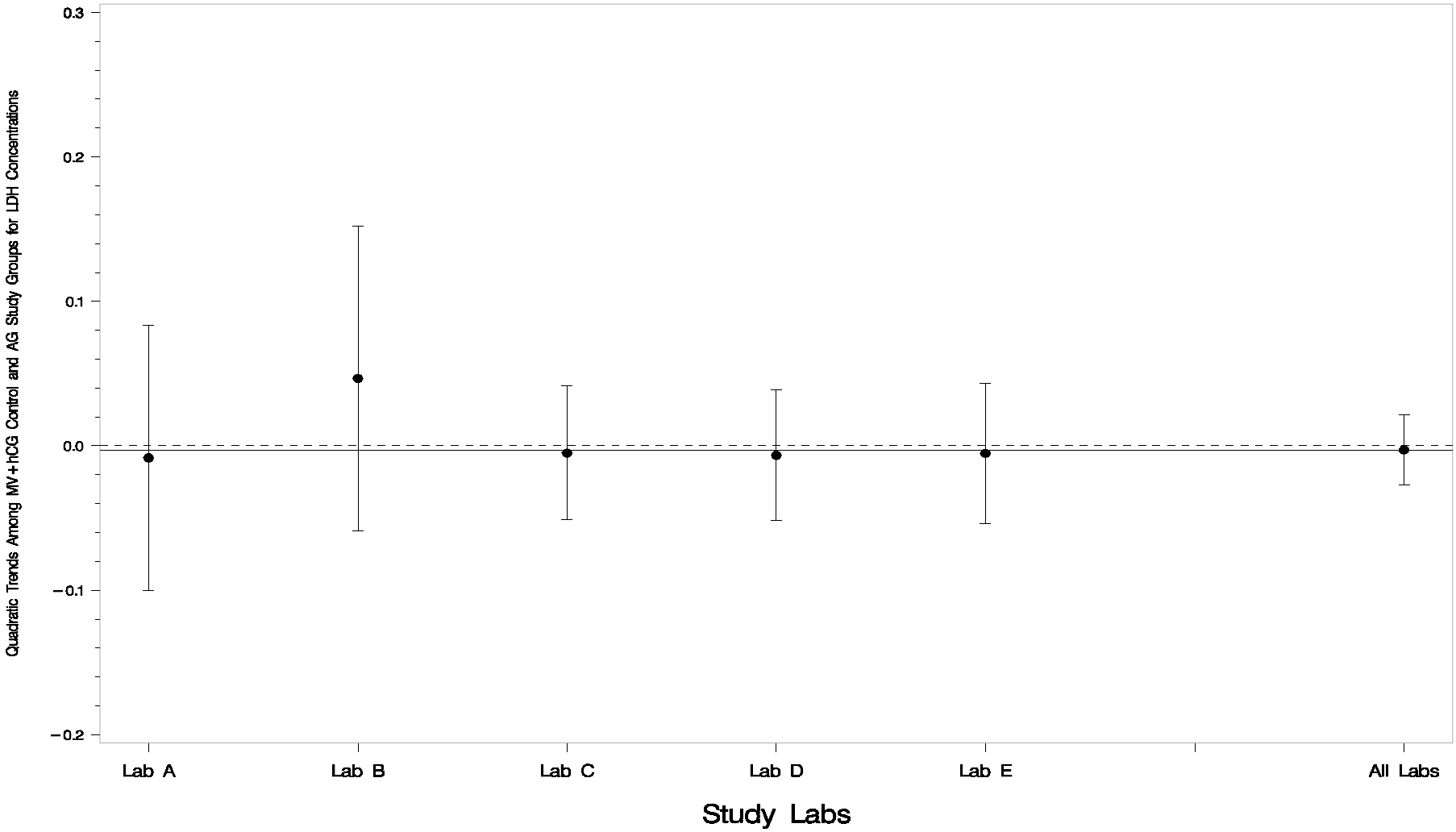

Figure 3.4-13. Quadratic Trends Among the M-V +hCG Control and the Three AG Study Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

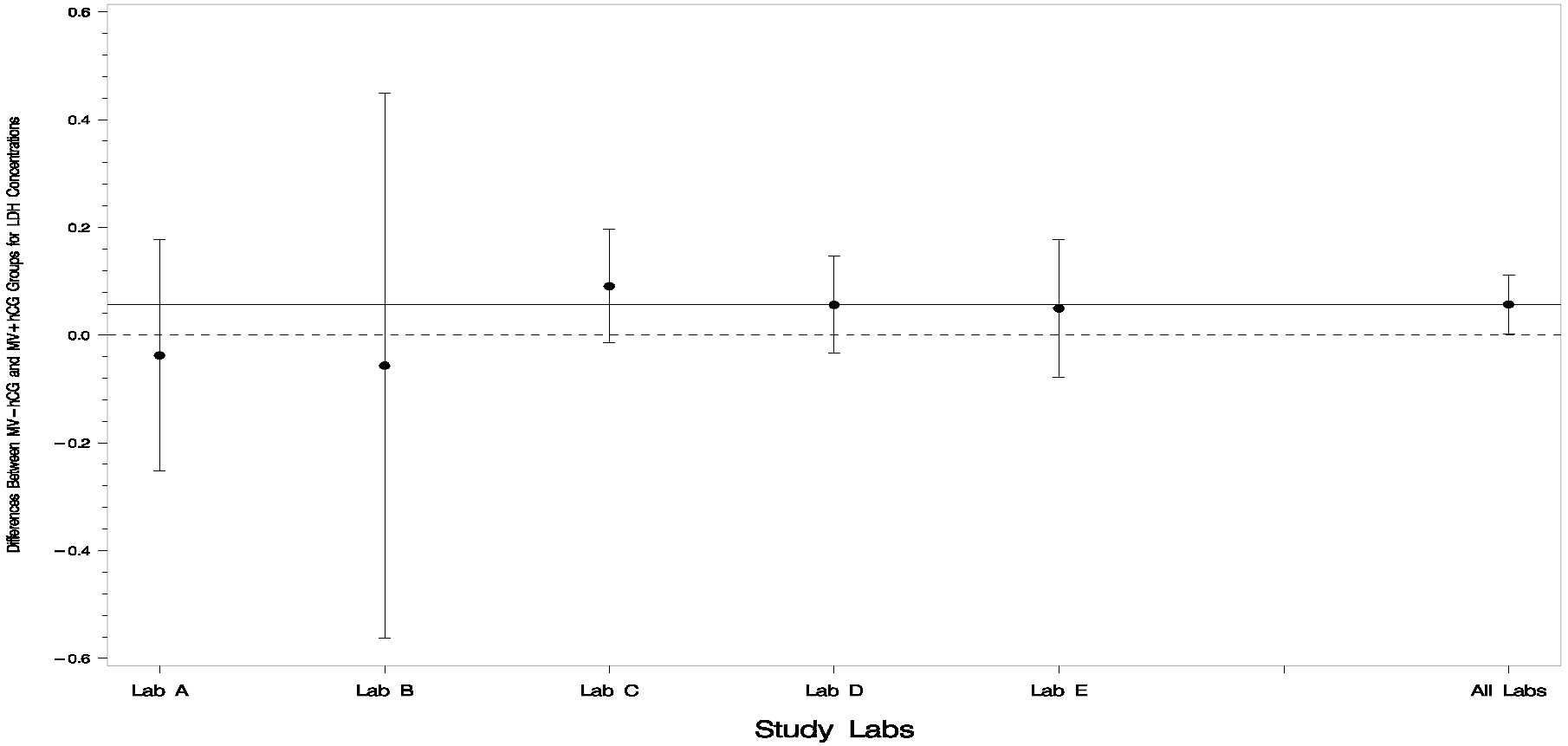

Figure 3.4-14. Differences Between M-V –hCG and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

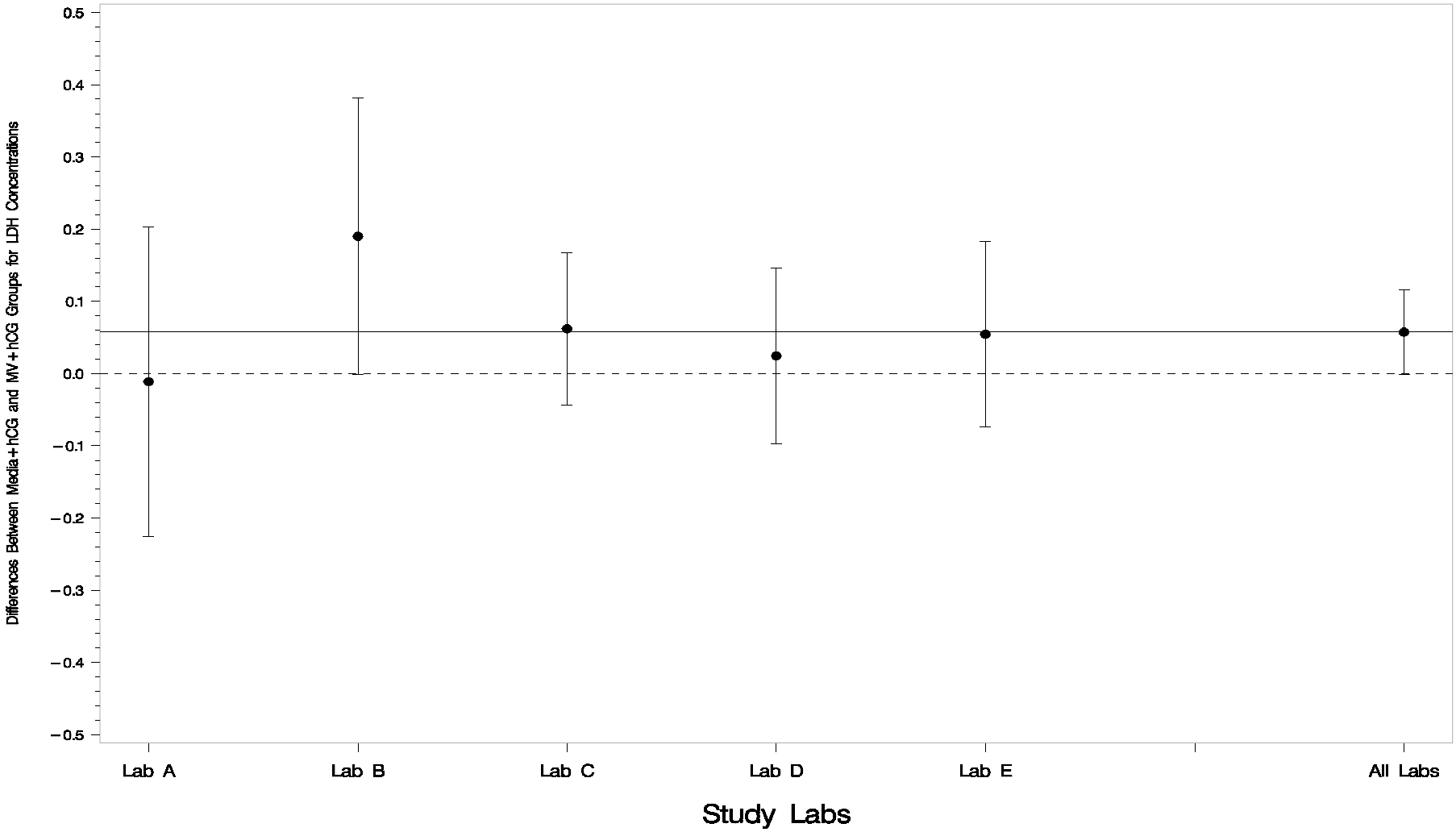

Figure 3.4-15. Differences Between Media+hCG and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

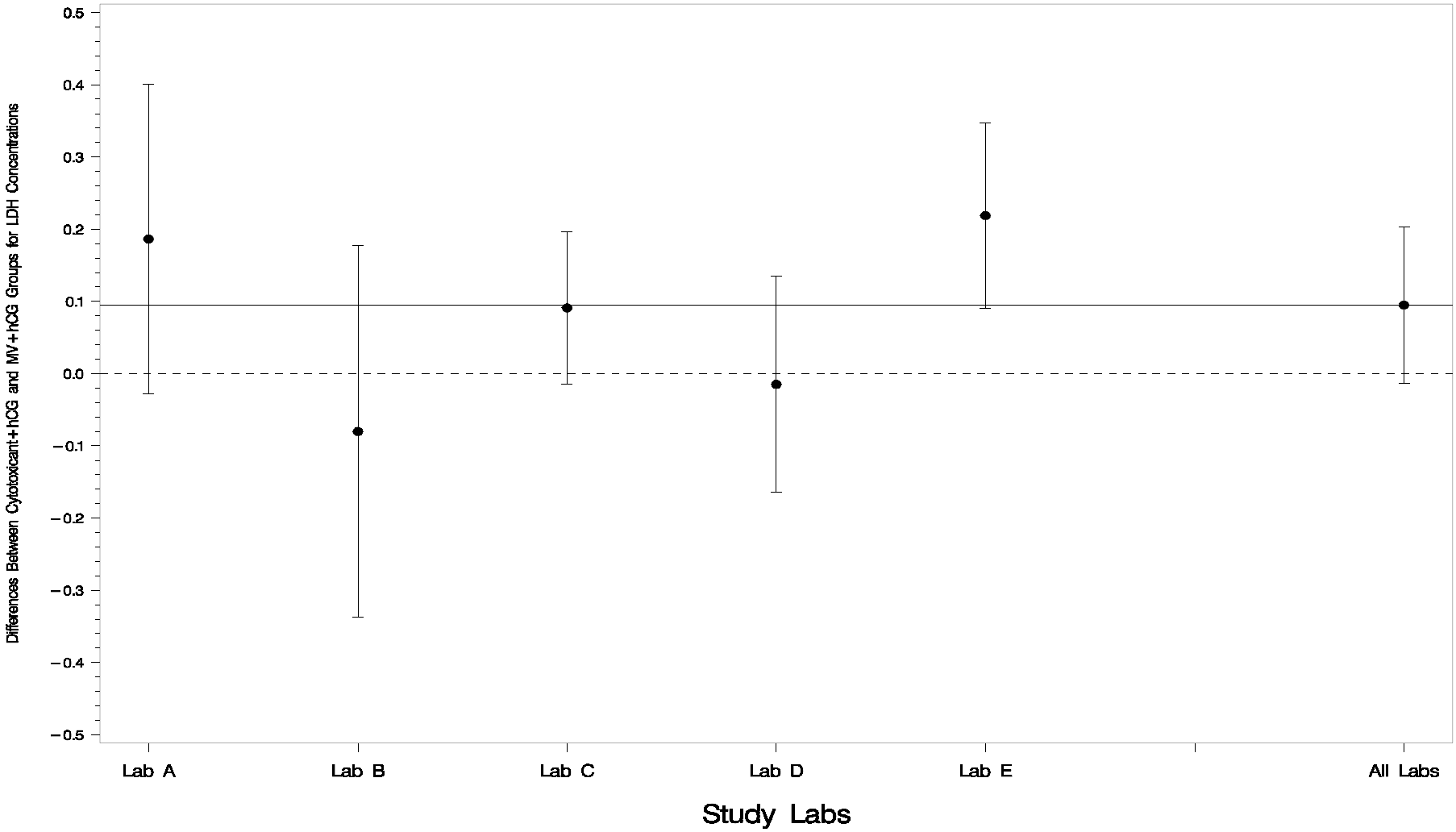

Figure 3.4-16. Differences Between Cytotoxicant+hCG and M-V +hCG Groups and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations, Across Laboratories and by Each Laboratory

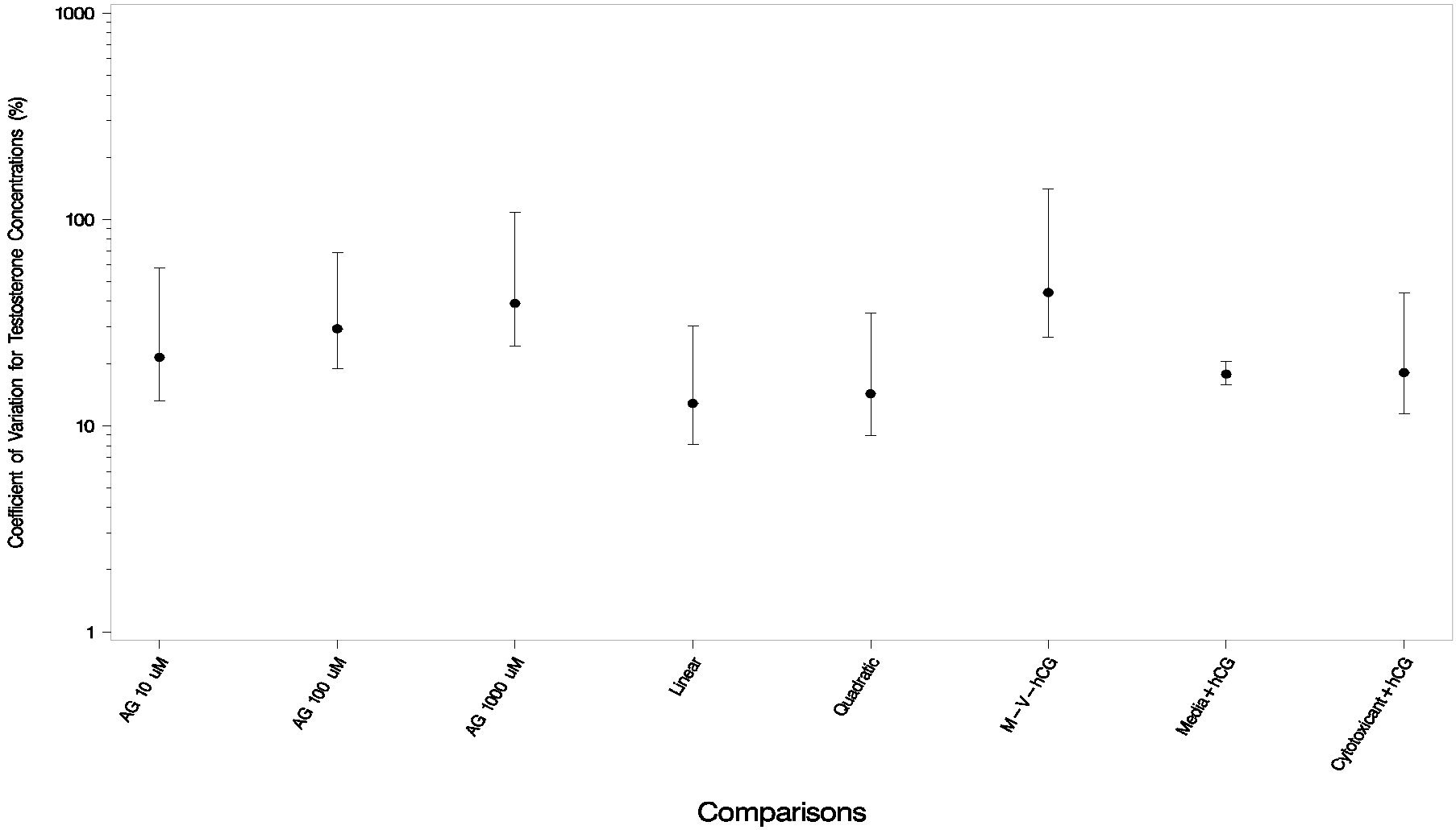

Figure 3.4-17. Coefficient of Variation and the Associated 95% CI of Natural Log Transformed Baseline Adjusted Testosterone Concentrations (%), Across All Laboratories for Each Comparison with the M-V +hCG Group

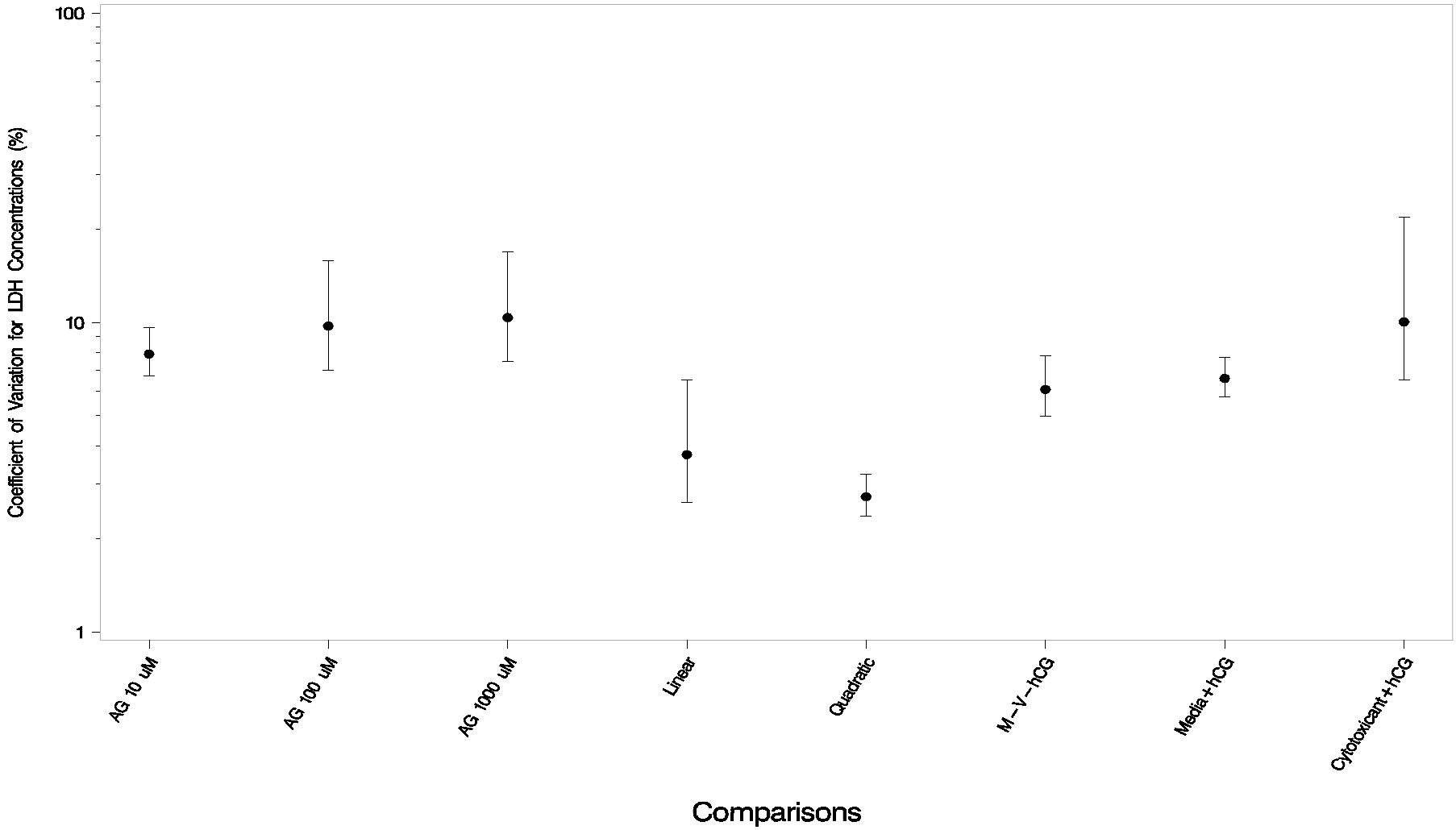

Figure 3.4-18. Coefficient of Variation and the Associated 95% CI of Natural Log Transformed Baseline Adjusted LDH Concentrations (%), Across All Laboratories for Each Comparison with the M-V +hCG Group

Table 3.4-7. Summary of CV% Values for Testosterone Across Laboratories for Individual Groups

Comparison	CV% Based on Logarithmic Mean and sd	CV% Based on Exponential Logarithmic Mean and sd[a]
M-V-hCG	36.1	33.1
M-V+hCG	47.0	43.8
Media +hCG	57.2	56.3
AG 10 µM	69.5	62.0
AG 100 µM	66.9	62.3
AG 1000 µM	31.8	28.4
EDS 1000 µM	58.5	56.5

a. Weighted values

Table 3.4-8. Summary of CV% Values for Testosterone Across Laboratories for Comparisons Among Groups

Comparison	CV% Based on Logarithmic Mean and sd	CV% Based on Exponential Logarithmic Mean and sd[a]
(M-V-hCG) vs (M-V+hCG)	44.15	39.32
(Media +hCG) vs (M-V+hCG)	17.77	18.17
AG 10 µM vs (M-V+hCG)	21.43	23.83
AG 100 µM vs (M-V+hCG)	29.48	27.82
AG 1000 µM vs (M-V+hCG)	39.10	36.42
EDS vs (M-V+hCG)	18.06	21.31
Linear Trend	12.81	12.72
Quadratic Trend	14.27	13.79

a. Weighted values

3.5 Supplemental Interlaboratory Comparisons

In addition to the preceding section, an additional interlaboratory analysis was made using the testosterone data. More specifically, this analysis involved using the testosterone summary results presented in Tables 3.1-1 through 3.1-3 for each of the seven endpoints determined in the study, as well as the ratios of these endpoints. In particular, the endpoints were:

- M-V Control -hCG
- M-V Control +hCG
- Media Control +hCG
- AG 10 μM
- AG 100 μM
- AG 1000 μM
- EDS 1000 μM

and the ratios evaluated were:

- M-V Control +hCG / M-V Control -hCG
- Media Control +hCG / M-V Control -hCG
- AG 10 μM / M-V Control +hCG
- AG 100 μM / M-V Control +hCG
- AG 1000 μM / M-V Control +hCG
- EDS 1000 μM / M-V Control +hCG

For the first two ratios, the outcome of the comparison was expressed as an "x-fold response" in testosterone concentration between the two groups, which was calculated using the following equation:

$$(T_{4\ Hr} - T_{0\ Hr})_{Control + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control - hCG}$$

The data were evaluated in this manner because the magnitude of the x-fold increase in testosterone concentration between the groups with and without hCG would be a measure of the sensitivity of the assay, i.e., the larger the x-fold increase, the more likely smaller effects of a given chemical with inhibitory activity would be able to be measured. A 10-fold increase was believed to be a very satisfactory performance criteria for the assay.

For the last four ratios, the outcome of the comparison was expressed as a "percent

inhibition" in testosterone concentration in the presence and absence of the inhibitor, which was calculated using the following equation:

$$\{[(T_{4\ Hr} - T_{0\ Hr})_{Inhibitor + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control + hCG}] \times 100\} - 100$$

The x-fold increases and % inhibition values by replicate and laboratory are presented in Tables 3.5-1 through 3.5-3 for replicates 1, 2, and 3, respectively. For each of the six ratios the summary values across laboratories and replicates were determined and compared. The statistical analysis results of these comparisons are presented in Tables 3.5-4 through 3.5-9. The information in these tables are organized as follows:

- the individual values are displayed by laboratory and by replicate within laboratory.
- the number of replicates, average values and variances are shown by laboratory.
- the overall average across laboratories is shown below the individual laboratory averages.
- a one-way analysis of variance (K=5 laboratories, n=3 replicates per laboratory) table is presented and summary values based on the analysis of variance are shown below the table.
- the overall variance within laboratories and degrees of freedom are shown.
- the laboratories are treated as a random effect and the variance between laboratories is shown.
- a 95% confidence on the overall average is given. The confidence interval incorporates laboratory-to-laboratory variation, as well as within laboratory variation.
- the 95% confidence intervals on each of the individual laboratory averages are given. These confidence intervals incorporate just within laboratory variation.

3.5.1 X-Fold Response for the Media Control + hCG and Media-Vehicle Control - hCG Groups

Testosterone concentrations were increased by the addition of hCG and the magnitude of this increase was evaluated by comparing the composite (4 Hrs) testosterone concentrations (after subtracting the baseline concentration) of the control groups in the presence and absence of hCG. The x-fold increase in testosterone concentration between the control groups with and without hCG ranged from 1.1 to 14.3 for the five laboratories. Most of the replicates and laboratories had less than 10-fold increases. A $\geq$10-fold increase was achieved for two of three replicates in two of the five laboratories (Labs C and E). The overall increase in testosterone concentration was 7.1-fold and the overall standard error was 1.4. Other statistical information concerning the x-fold comparison results for these two groups across all laboratories is summarized in Table 3.5-4.

3.5.2 X-Fold Response for the Media-Vehicle Control + hCG and Media-Vehicle Control - hCG Groups

Testosterone concentrations were increased by the addition of hCG and the magnitude of this increase was evaluated by comparing the composite (4 Hrs) testosterone concentrations (after subtracting the baseline concentration) of the control groups in the presence and absence of hCG. The x-fold increase in testosterone concentration between the control groups with and without hCG ranged from 2.3 to 15.6 for the five laboratories. Most of the replicates and laboratories had less than 10-fold increases. A $\geq$10-fold increase was achieved for two of three replicates in two of the five laboratories (Labs C and E). The overall increase in testosterone concentration was 7.5-fold and the overall standard error was 1.5. Other statistical information concerning the x-fold comparison results for these two groups across all laboratories is summarized in Table 3.5-5.

Table 3.5-1. Interlaboratory Comparison of the Testosterone Response Expressed as an "x-fold" Change Between Controls and "% inhibition" With and Without an Inhibitor - Replicate 1 Data Set

Group	Amino-glutethimide (AG) Conc	hCG	Lab A			Lab D		
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]
			0 Hr (Baseline)	4 Hr (Comp.)		0 Hr (Baseline)	4 Hr (Comp.)	
Media-Vehicle Control	0	-	0.02	0.54	--	0.05	1.16	--
Media Control	0	+	0.01	2.89	5.5x	0.04	1.27	1.1x
Media-Vehicle Control	0	+	0.02	3.52	6.7x	0.06	2.12	2.3x
AG	10 uM	+	0.02	2.63	- 25.4%	0.05	0.7	-75.0%
	100 uM	+	0.02	0.34	- 90.9%	0.05	0.29	-90.8%
	1000 uM	+	0.02	0.09	- 98.0%	0.04	0.16	-95.4%
EDS (1000 uM)	0	+	0.02	0.7	- 80.6%	0.04	0.4	-86.2%

a. Control Response as an "x- fold" increase = $(T_{4\ Hr} - T_{0\ Hr})_{Control + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control - hCG}$

b. Inhibitory Response as a % of Control = $\{[(T_{4\ Hr} - T_{0\ Hr})_{Inhibitor + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control + hCG}] \times 100\} - 100$

Table 3.5-1. Interlaboratory Comparison of the Testosterone Response Expressed as an "x-fold" Change Between Controls and "% inhibition" With and Without an Inhibitor - Replicate 1 Data Set (continued)

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab E			Lab B		
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]
			0 Hr (Baseline)	4 Hr (Comp.)		0 Hr (Baseline)	4 Hr (Comp.)	
Media-Vehicle Control	0	-	0.02	0.24	--	0.08	0.84	--
Media Control	0	+	0.03	2.11	9.5x	0.09	3.26	4.2x
Media-Vehicle Control	0	+	0.03	3.17	14.3x	0.1	3.33	4.3x
AG	10 uM	+	0.03	1.93	- 39.5%	0.09	2.55	-23.8%
	100 uM	+	0.02	0.27	- 92.0%	0.1	0.95	-73.7%
	1000 uM	+	0.03	0.11	- 97.5%	0.09	0.32	-92.9%
EDS (1000 uM)	0	+	0.03	0.4	- 88.2%	0.09	0.77	-78.9%

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab C			Replicate 1 Overall Response (x-fold) [a] (% I) [b] mean ± SD
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	
			0 Hr (Baseline)	4 Hr (Comp.)		
Media-Vehicle Control	0	-	0.06	0.76	--	--
Media Control	0	+	0.09	8.45	11.9x	7.8 ± 3.6 (n=4)
Media-Vehicle Control	0	+	0.09	7.57	10.7x	9.0 ± 4.4 (n=4)
AG	10 uM	+	0.08	5.6	- 26.2%	-38.0 ± 21.6 (n=5)
	100 uM	+	0.08	1.87	- 76.1%	-84.7 ± 9.0 (n=5)
	1000 uM	+	0.09	0.31	- 97.1%	-96.2 ± 2.1 (n=5)
EDS (1000 uM)	0	+	0.08	2.2	- 71.7%	-81.1 ± 6.5 (n=5)

Table 3.5-2. Interlaboratory Comparison of the Testosterone Response Expressed as an "x-fold" Change Between Controls and "% inhibition" With and Without an Inhibitor - Replicate 2 Data Set

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab A			Lab D		
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]
			0 Hr (Baseline)	4 Hr (Comp.)		0 Hr (Baseline)	4 Hr (Comp.)	
Media-Vehicle Control	0	-	0.02	0.48	--	0.03	0.68	--
Media Control	0	+	0.03	4.25	9.2x	0.04	3.13	4.8x
Media-Vehicle Control	0	+	0.02	3.48	7.5x	0.03	3.58	5.5x
AG	10 uM	+	0.02	1.31	- 62.7%	0.04	1.47	-59.7%
	100 uM	+	0.02	0.42	- 88.4%	0.04	0.52	-86.5%
	1000 uM	+	0.02	0.13	- 96.8%	0.04	0.16	-96.6%
EDS (1000 uM)	0	+	0.02	0.91	- 74.3%	0.04	1.27	-65.4%

a. Control Response as an "x- fold" increase = $(T_{4\ Hr} - T_{0\ Hr})_{Control + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control - hCG}$

b. Inhibitory Response as a % of Control = $\{[(T_{4\ Hr} - T_{0\ Hr})_{Inhibitor + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control + hCG}] \times 100\} - 100$

Table 3.5-2. Interlaboratory Comparison of the Testosterone Response Expressed as an "x-fold" Change Between Controls and "% inhibition" With and Without an Inhibitor - Replicate 2 Data Set (Continued)

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab E			Lab B		
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]
			0 Hr (Baseline)	4 Hr (Comp.)		0 Hr (Baseline)	4 Hr (Comp.)	
Media-Vehicle Control	0	-	0.06	0.57	--	0.08	1.56	--
Media Control	0	+	0.05	2.53	4.9x	0.09	6.17	4.1x
Media-Vehicle Control	0	+	0.06	2.29	4.4x	0.09	5.61	3.7x
AG	10 uM	+	0.06	1.61	- 30.5%	0.1	3.26	-42.8%
	100 uM	+	0.04	0.32	- 87.4%	0.09	0.78	-87.5%
	1000 uM	+	0.05	0.21	- 92.8%	0.08	0.27	-96.6%
EDS (1000 uM)	0	+	0.05	0.59	- 75.8%	0.06	1.05	-80.3%

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab C			Overall Response (x-fold) [a] (% I) [b] mean ± SD
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	
			0 Hr (Baseline)	4 Hr (Comp.)		
Media-Vehicle Control	0	-	0.04	0.82	--	--
Media Control	0	+	0.05	7.55	9.6x	6.4 ± 2.7 (n=4)
Media-Vehicle Control	0	+	0.05	6.87	8.7x	6.0 ± 2.0 (n=4)
AG	10 uM	+	0.05	3.93	- 43.1%	-49.0 ± 11.4 (n=5)
	100 uM	+	0.04	0.89	- 87.5%	-87.7 ± 0.9 (n=5)
	1000 uM	+	0.05	0.81	- 98.1%	-96.3 ± 1.7 (n=5)
EDS (1000 uM)	0	+	0.07	1.42	- 80.2%	-75.7 ± 6.2 (n=5)

Table 3.5-3. Interlaboratory Comparison of the Testosterone Response Expressed as an "x-fold" Change Between Controls and "% inhibition" With and Without an Inhibitor - Replicate 3 Data Set

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab A			Lab D		
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]
			0 Hr (Baseline)	4 Hr (Comp.)		0 Hr (Baseline)	4 Hr (Comp.)	
Media-Vehicle Control	0	-	0.08	0.71	--	0.02	0.39	--
Media Control	0	+	0.06	5.35	8.4x	0.02	1.86	5.0x
Media-Vehicle Control	0	+	0.06	5.14	8.1x	0.02	1.71	4.6x
AG	10 uM	+	0.06	2.33	- 55.3%	0.02	0.84	-51.5%
	100 uM	+	0.07	0.66	- 88.4%	0.03	0.22	-75.1%
	1000 uM	+	0.06	0.22	- 96.9%	0.03	0.1	-95.9%
EDS (1000 uM)	0	+	0.06	1.2	- 77.6%	0.03	0.52	-71.0%

a. Control Response as an "x- fold" increase = $(T_{4\ Hr} - T_{0\ Hr})_{Control + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control - hCG}$

b. Inhibitory Response as a % of Control = $\{[(T_{4\ Hr} - T_{0\ Hr})_{Inhibitor + hCG} \div (T_{4\ Hr} - T_{0\ Hr})_{M\text{-}V\ Control + hCG}] \times 100\} - 100$

Table 3.5-3. Interlaboratory Comparison of the Testosterone Response Expressed as an "x-fold" Change Between Controls and "% inhibition" With and Without an Inhibitor - Replicate 3 Data Set (Continued)

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab E			Lab B		
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]
			0 Hr (Baseline)	4 Hr (Comp.)		0 Hr (Baseline)	4 Hr (Comp.)	
Media-Vehicle Control	0	-	0.04	0.3	--	0.06	0.68	--
Media Control	0	+	0.04	2.14	8.1x	0.05	3.81	6.1x
Media-Vehicle Control	0	+	0.04	2.74	10.4x	0.04	3.32	5.3x
AG	10 uM	+	0.03	1.05	- 62.2%	0.04	1.96	-41.5%
	100 uM	+	0.04	0.46	- 84.4%	0.05	0.93	-73.2%
	1000 uM	+	0.03	0.13	- 96.3%	0.05	0.25	-93.9%
EDS (1000 uM)	0	+	0.03	0.42	- 85.6%	0.05	1.33	-61.0%

Group	Amino-glutethimide (AG) Conc	hCG Present	Lab C			Overall Response (x-fold) [a] (% I) [b] mean ± SD
			Testosterone Conc (ng/mg)		Response (x-fold) [a] (% I) [b]	
			0 Hr (Baseline)	4 Hr (Comp.)		
Media-Vehicle Control	0	-	0.06	0.65	--	--
Media Control	0	+	0.05	8.46	14.3x	8.4 ± 3.6 (n=4)
Media-Vehicle Control	0	+	0.05	9.23	15.6x	8.8 ± 4.5 (n=4)
AG	10 uM	+	0.04	5.92	- 35.9%	-49.1 ± 10.2 (n=5)
	100 uM	+	0.05	1.48	- 84.4%	-81.1 ± 6.6 (n=5)
	1000 uM	+	0.05	0.2	- 98.4%	-96.1 ± 1.7 (n=5)
EDS (1000 uM)	0	+	0.04	2.65	- 71.6%	-73.4 ± 9.1 (n=5)

Table 3.5-4. X-Fold Comparison Between Media Control + hCG and Media-Vehicle Control - hCG

Replicate	Lab A	Lab D	Lab E	Lab B	Lab C
1	5.5	1.1	9.5	4.2	11.9
2	9.2	4.8	4.9	4.1	9.6
3	8.4	5.0	8.1	6.1	14.3

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	3	23.1	7.70	3.79
Column 2	3	10.9	3.63	4.82
Column 3	3	22.5	7.50	5.56
Column 4	3	14.4	4.80	1.27
Column 5	3	35.8	11.93	5.52
			7.11	4.19

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	123.564	4	30.891	7.36669316	0.004943	3.47805
Within Groups	41.933	10	4.193333			
Total	165.497	14				

n per Group	3		t(0.975,4)=	2.776		
K Groups	5		t(0.975,10)=	2.228		
Var within groups	4.19	df	10			
Var betwn groups	8.90					
					95% CI	
Overall Average	7.11	Overall Std Err	1.44	df = 4	3.129604	11.09706
Lab A Avg	7.70	W/I Lab Std Err	1.18	df = 10	5.07	10.33
Lab D Avg	3.63	W/I Lab Std Err	1.18	df = 10	1.00	6.27
Lab E Avg	7.33	W/I Lab Std Err	1.18	df = 10	4.70	9.97
Lab B Avg	4.80	W/I Lab Std Err	1.18	df = 10	2.17	7.43
Lab C Avg	11.93	W/I Lab Std Err	1.18	df = 10	9.30	14.57

Table 3.5-5. X-Fold Comparison Between M-V Control + hCG and M-V Control - hCG

Replicate	Lab A	Lab D	Lab E	Lab B	Lab C
1	6.7	2.3	14.3	4.3	10.7
2	7.5	5.5	4.4	3.7	8.7
3	8.1	4.6	10.4	5.3	15.6

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	3	22.3	7.43	0.49
Column 2	3	12.4	4.13	2.72
Column 3	3	29.1	9.70	24.87
Column 4	3	13.3	4.43	0.65
Column 5	3	35	11.67	12.60
			7.47	8.27

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	128.823	4	32.20567	3.89490446	0.036939	3.47805
Within Groups	82.687	10	8.268667			
Total	211.509	14				

					95% CI	
n per Group	3		t(0.975,4)=	2.776		
K Groups	5		t(0.975,10)=	2.228		
Var within groups	8.27	df	10			
Var betwn groups	7.98					
Overall Average	7.47	Overall Std Err	1.47	df =	4	3.41 11.54
Lab A Avg	7.43	W/I Lab Std Err	1.66	df =	10	3.73 11.13
Lab D Avg	4.13	W/I Lab Std Err	1.66	df =	10	0.43 7.83
Lab E Avg	9.70	W/I Lab Std Err	1.66	df =	10	6.00 13.40
Lab B Avg	4.43	W/I Lab Std Err	1.66	df =	10	0.73 8.13
Lab C Avg	11.67	W/I Lab Std Err	1.66	df =	10	7.97 15.37

3.5.3 Percent Inhibition of Testosterone Production by Aminoglutethimide (AG)

Testosterone concentrations were decreased in a concentration-dependent manner in the presence of increasing concentrations of AG ranging from 10 to 1000 µM for all laboratories when compared to the testosterone concentrations produced by the M-V Control +hCG group. The statistical information concerning the % inhibition results for the 10, 100, and 1000 µM AG groups by replicate and laboratory is summarized in Tables 3.5-6 through 3.5-8, respectively.

At a concentration of 10 µM, AG decreased the testosterone concentration. The percent inhibition values ranged from -23.8 to -75.0 for all replicates and laboratories. Most laboratories reported a decrease in the testosterone concentration relative to the M-V Control +hCG group of approximately 50 percent or less at a concentration of 10 µM AG. The overall percent inhibition was -45.0 percent with a standard error of 4.9 percent. There were no significant differences in the source of variation between or within groups.

At a concentration of 100 µM, AG decreased the testosterone concentration to an even greater extent than was observed at 10 µM. The percent inhibition values ranged from -73.2 to -92.0 for all replicates and laboratories. Most laboratories reported a decrease in the testosterone concentration relative to the M-V Control +hCG group of approximately 80 to 90 percent at this concentration of AG. The overall percent inhibition was -84.4 percent with a standard error of 2.0 percent. There were no significant differences in the source of variation between or within groups.

At a concentration of 1000 µM, AG decreased the testosterone concentration to an even greater extent than was observed at 100 µM. The percent inhibition values ranged from -92.8 to -98.4 for all replicates and laboratories. Most laboratories reported a decrease in the testosterone concentration relative to the M-V Control+hCG group that was greater than 95 percent at this concentration of AG. The overall percent inhibition was -96.2 percent with a standard error of 0.6 percent. There were no significant differences in the source of variation between or within groups.

3.5.4 Percent Inhibition of Testosterone Production by EDS

Testosterone concentrations were decreased in the presence of EDS at a concentration of 1000 µM for all laboratories when compared to the testosterone concentrations produced by the M-V Control + hCG group. The statistical information concerning the % inhibition results is summarized in Table 3.5-9.

At a concentration of 1000 µM, EDS decreased the testosterone concentration (relative to the M-V +hCG group). The percent inhibition values ranged from -61.0 to -88.2 for all replicates and laboratories. The overall percent inhibition was -76.6 percent with a standard error of 2.0 percent. There were no significant differences in the source of variation between or within groups.

Table 3.5-6. Percent Inhibition in Testosterone Concentration by Aminoglutethimide at 10 uM

Replicate	Lab A	Lab D	Lab E	Lab B	Lab C
1	-25.4	-75.0	-39.5	-23.8	-26.2
2	-62.7	-59.7	-30.5	-42.8	-43.1
3	-55.3	-51.5	-62.2	-41.5	-35.9

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	3	-143.4	-47.80	390.01
Column 2	3	-186.2	-62.07	142.26
Column 3	3	-132.2	-44.07	266.86
Column 4	3	-108.1	-36.03	112.66
Column 5	3	-105.2	-35.07	71.92
			-45.01	196.74

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	1437.163	4	359.2907	1.82617741	0.200458	3.47805
Within Groups	1967.447	10	196.7447			
Total	3404.609	14				

					95% CI	
n per Group	3		t(0.975,4)=	2.776		
K Groups	5		t(0.975,10)=	2.228		
Var within groups	196.74	df	10			
Var betwn groups	54.18					
Overall Average	-45.01	Overall Std Err	4.89	df = 4	-58.59	-31.42
Lab A Avg	-47.80	W/I Lab Std Err	8.10	df = 10	-65.84	-29.76
Lab D Avg	-62.07	W/I Lab Std Err	8.10	df = 10	-80.11	-44.02
Lab E Avg	-44.07	W/I Lab Std Err	8.10	df = 10	-62.11	-26.02
Lab B Avg	-36.03	W/I Lab Std Err	8.10	df = 10	-54.08	-17.99
Lab C Avg	-35.07	W/I Lab Std Err	8.10	df = 10	-53.11	-17.02

Table 3.5-7. Percent Inhibition in Testosterone Concentration by Aminoglutethimide at 100 uM

Replicate	Lab A	Lab D	Lab E	Lab B	Lab C
1	-90.9	-90.8	-92.0	-73.7	-76.1
2	-88.4	-86.5	-87.4	-87.5	-87.5
3	-88.4	-75.1	-84.4	-73.2	-84.4

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	3	-267.7	-89.23	2.08
Column 2	3	-252.4	-84.13	65.82
Column 3	3	-263.8	-87.93	14.65
Column 4	3	-234.4	-78.13	65.86
Column 5	3	-248	-82.67	34.74
			-84.42	36.63

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	234.571	4	58.64267	1.600800728	0.248568	3.47805
Within Groups	366.333	10	36.63333			
Total	600.904	14				

n per Group	3		t(0.975,4)=	2.776	
K Groups	5		t(0.975,10)=	2.228	
Var within groups	36.63df		10		
Var betwn groups	7.34				
					95% CI
Overall Average	-84.42	Overall Std Err	1.98df =	4	-89.91 -78.93
Lab A Avg	-89.23	W/I Lab Std Err	3.49df =	10	-97.02 -81.45
Lab D Avg	-84.13	W/I Lab Std Err	3.49df =	10	-91.92 -76.35
Lab E Avg	-87.93	W/I Lab Std Err	3.49df =	10	-95.72 -80.15
Lab B Avg	-78.13	W/I Lab Std Err	3.49df =	10	-85.92 -70.35
Lab C Avg	-82.67	W/I Lab Std Err	3.49df =	10	-90.45 -74.88

Table 3.5-8. Percent Inhibition in Testosterone Concentration by Aminoglutethimide at 1000 uM

Replicate	Lab A	Lab D	Lab E	Lab B	Lab C
1	-98.0	-95.4	-97.5	-92.9	-97.1
2	-96.8	-96.6	-92.8	-96.6	-98.1
3	-96.9	-95.9	-96.3	-93.9	-98.4

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	3	-291.7	-97.23	0.44
Column 2	3	-287.9	-95.97	0.36
Column 3	3	-286.6	-95.53	5.96
Column 4	3	-283.4	-94.47	3.66
Column 5	3	-293.6	-97.87	0.46
			-96.21	2.18

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	22.044	4	5.511	2.52875497	0.106718	3.47805
Within Groups	21.793	10	2.179333			
Total	43.837	14				

					95% CI	
n per Group	3		t(0.975,4)=	2.776		
K Groups	5		t(0.975,10)=	2.228		
Var within groups	2.18	df	10			
Var betwn groups	1.11					
Overall Average	-96.21	Overall Std Err	0.61	df =	-97.90	-94.53
Lab A Avg	-97.23	W/I Lab Std Err	0.85	df = 10	-99.13	-95.33
Lab D Avg	-95.97	W/I Lab Std Err	0.85	df = 10	-97.87	-94.07
Lab E Avg	-95.53	W/I Lab Std Err	0.85	df = 10	-97.43	-93.63
Lab B Avg	-94.47	W/I Lab Std Err	0.85	df = 10	-96.37	-92.57
Lab C Avg	-97.87	W/I Lab Std Err	0.85	df = 10	-99.77	-95.97

Table 3.5-9. Percent Inhibition in Testosterone Concentration by EDS at 1000 uM

Replicate	RTI	Battelle	Southern	Toxikon	WIL
1	-80.6	-86.2	-88.2	-78.9	-71.7
2	-74.3	-65.4	-75.8	-80.3	-80.2
3	-77.6	-71.0	-85.6	-61.0	-71.6

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	3	-232.5	-77.50	9.93
Column 2	3	-222.6	-74.20	115.84
Column 3	3	-249.6	-83.20	42.76
Column 4	3	-220.2	-73.40	115.81
Column 5	3	-223.5	-74.50	24.37
			-76.56	61.74

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	194.316	4	48.579	0.786806388	0.559298	3.47805
Within Groups	617.42	10	61.742			
Total	811.736	14				

n per Group	3		t(0.975,4)=	2.776		
K Groups	5		t(0.975,10)=	2.228		
Var within groups	61.74	df	10			
Var betwn groups	0.00					
					95% CI	
Overall Average	-76.56	Overall Std Err	2.03	df = 4	-82.19	-70.93
Lab A Avg	-77.50	W/I Lab Std Err	4.54	df = 10	-87.61	-67.39
Lab B Avg	-74.20	W/I Lab Std Err	4.54	df = 10	-84.31	-64.09
Lab E Avg	-83.20	W/I Lab Std Err	4.54	df = 10	-93.31	-73.09
Lab B Avg	-73.40	W/I Lab Std Err	4.54	df = 10	-83.51	-63.29
Lab C Avg	-74.50	W/I Lab Std Err	4.54	df = 10	-84.61	-64.39

3.6 Laboratory Procedural Findings

Specific steps of the assay and the time taken to perform these steps are considered very important because they were believed to affect the viability of the testes and fragments as well the intra- and interlaboratory variability. The steps of special interest include testis isolation, weight, and collection; testis slicing; media sample collection; and assay start-to-finish time. In addition, observations made by the laboratory staff, during assay conduct and the Work Assignment Leader (WAL), while monitoring the laboratories, are described as these too were believed to impact the outcome of the task and provide further understanding to the variability of the assay.

3.6.1 Testis Isolation, Weight, and Collection

The right testis weights (mean ± sd, n = 9), by laboratory, were 1.6 ± 0.1, 1.6 ± 0.2, 1.6 ± 0.2, 1.8 ± 0.2, and 1.7 ± 0.1, grams for Labs A through E. All testes were within the protocol required specification, i.e., >1000 mg.

The span of time from immediately after animal termination until a given testis was isolated, removed, and immersed in cold buffered media was considered an important step in assuring tissue viability. The results are summarized in Table 3.6-1. Most laboratories averaged 5 minutes or less to perform this step, although two of the laboratories did not record sufficient information to be able to determine the elapsed time for any (Lab B) or two (Lab C) of the three replicates.

Table 3.6-1. Elapsed Time from Time of Death to Testis Immersion in Cold Buffered Media

Trial	Elapsed Time (min)[a]				
	Lab A	Lab B[b]	Lab C[c]	Lab D	Lab E
Replicate 1	6.0 ± 1.4	ND	3.0 ± 1.0	2.3 ± 2.1	3.0 ± 1.0
Replicate 2	4.5 ± 0.7	ND	ND	1.0 ± 1.0	2.3 ± 0.5
Replicate 3	4.8 ± 0.4	ND	ND	2.0 ± 1.0	3.0 ± 1.7
Within Lab Mean ± sd	5.1 ± 1.0	ND	ND	1.8 ± 1.4	2.7 ± 0.5

a. Mean ± sd, n = 3.

b. Lab B did not determine the elapsed time to perform this step but reported an estimated elapsed time of 5 minutes.

c. Lab C did not determine the elapsed time to perform this step for replicates 2 and 3 but estimated the elapsed time for these replicates to be similar to the replicate 1 time.

3.6.2. Testis Slicing

The elapsed time from when a given testis was removed from the cold buffered media until the last (7^{th}) fragment for that testis was sliced, collected, and placed in freshly gassed media is considered an important step in assuring tissue viability. The results are summarized in Table 3.6-2. Two of the laboratories processed the testis into fragments in 6 minutes or less (Labs C and D), whereas 10 minutes or more occurred for the other labs. It is important to note that the table below provides information about the greatest length of time that the testis was out of the cold buffered media until the freshly sliced fragment was placed in freshly gassed media. However, it does not provide information about the delay in processing a testis into fragments. For example, one laboratory (Lab D) reported this time delay and the overall elapsed time was 31 ± 11 min, thereby indicating that while the testis slicing process, once started, proceeded within several minutes, the time that the testis remained in the cold buffered media before slicing commenced exceeded half an hour for all three replicates for this laboratory.

Table 3.6-2. Testis to Fragment Processing Time

Trial	Elapsed Time (min)[a]				
	Lab A[b]	Lab B[c]	Lab C	Lab D	Lab E
Replicate 1	11.7	ND	6.0 ± 1.0	7.0 ± 1.7	15.0 ± 3.0
Replicate 2	10.3	ND	5.7 ± 0.6	6.0 ± 2.0	21.7 ± 3.5
Replicate 3	12.7	ND	3.7 ± 0.6	5.0 ± 1.0	21.3 ± 4.7
Within Lab Mean ± sd	11.6 ± 1.2	ND	5.1 ± 1.3	6.0 ± 2.0	19.4 ± 2.1

a. Mean ± sd, n = 3.

b. Tabled values were estimated by dividing the laboratory reported values by three in order to obtain a time period for comparison with the other laboratory values. The laboratory reported value was the time taken to slice 3 testis and distributing the 21 fragments into test tubes.

c. Not determined by laboratory but estimated in their report to be 10 - 15 minutes.

3.6.3 Media Sample Collection

The removal time of the fragments from the incubator, through sample collection, until the fragments were returned to the incubator is considered an important step in assuring tissue viability. The results for the 1- and 3-hour collections are summarized in Table 3.6-3. Four of five laboratories completed this step in approximately 20 minutes or less, while one laboratory (Lab D) averaged approximately 30 minutes at each collection time point.

Table 3.6-3. Media Sample Collection Processing Time

Trial	Elapsed Time (min)				
	Lab A	Lab B	Lab C	Lab D	Lab E
1-Hour Sample					
Replicate 1	12	NR[a]	19	32	22
Replicate 2	13	NR	17	30	26
Replicate 3	12	NR	24	35	21
1-Hr Mean ± sd	12 ± 1	19 ± 3	20 ± 4	32 ± 3	23 ± 3
3-Hour Sample					
Replicate 1	13	NR	19	30	23
Replicate 2	13	NR	21	40	21
Replicate 3	12	NR	22	30	23
3-Hr Mean ± sd	13 ± 1	21 ± 2	21 ± 2	33 ± 6	22 ± 1
Within Lab Mean ± sd	13 ± 1	20[b]	20 ± 3	33 ± 4	23 ± 2

a. Not reported.

b. Average of the 1- and 3-hour mean values.

3.6.4 Total Assay Execution Time

The time taken from start to finish (overall assay time from time of animal death of the first animal until the collection of the final four hour sample) is an important consideration in assuring tissue viability. The results are summarized in Table 3.6-4. For a given laboratory, the assay execution time generally decreased with experience. Most laboratories performed the assay as designed in approximately 7 hours or less, while one laboratory (Lab D) averaged an additional 1.5 hours.

Table 3.6-4. Total Assay Execution Time

Trial	Elapsed Time (hours)				
	Lab A	Lab B	Lab C	Lab D	Lab E
Replicate 1	6.6	7.6	7	9.3	7.1
Replicate 2	6.3	6.9	6.9	7.9	7
Replicate 3	6.5	6.7	6.8	8.3	6.8
Within Lab Mean ± sd	6.4 ± 0.1	7.1 ± 0.5	6.9 ± 0.1	8.5 ± 0.7	7.0 ± 0.2

3.6.5 Laboratory Procedural Deviations

This section summarizes deviations from the protocol that had minor to severe impacts on the assay. In those instances when the impact was determined to have a severe impact, then the laboratory was required to repeat the replicate. Two of the five laboratories were required to repeat one replicate. No laboratory had to repeat more than one replicate.

One laboratory did not use the correct number of rats or testis to obtain fragments. More specifically, this laboratory used the right testis isolated from two rats (instead of three) and, from these testes, distributed 12 fragments from the first rat testis into all cells of the three control groups and low concentration AG group, and 9 fragments from the second rat testis into all cells of the mid and low AG groups and cytotoxicant group (instead of distributing one fragment from each rat testis into each group). This deviation was considered to have a severe impact on the replicate study results. There was no obvious explanation for the procedural deviation other than technician error. This deviation was discovered upon review of the first replicate data set, which was required of the laboratories before they were approved to proceed to the second and third replicate studies.

One laboratory did not collect the required samples. During the conduct of the second replicate study, this laboratory collected the wash sample after equilibration but instead of adding blank media for collection of a baseline sample, it added the group-specific media and proceeded with the one hour incubation period. This deviation was considered to have a severe impact on the replicate study results. There was no obvious explanation other than technician error. This deviation was identified by the laboratory technician and reported by the study director.

One laboratory did not perform the pilot baseline/positive control study in a sufficiently acceptable manner to allow approval to proceed to the formal task. This laboratory conducted the pilot study using 12 mL test tubes (instead of the 9 mL, 13 x 100 mm test tube), set the incubation temperature at 37° C (instead of 36° C), oriented the tubes in the incubator in a

vertical position during shaking (instead of horizontal), and used an animal well beyond the 15 week-old age limit for testing. These findings were discovered upon review of the pilot study results. These deviations were believed to occur because, even though the laboratory personnel were given the optimized assay specifications, the staff made changes that they did not consider sufficiently different to make any difference.

Not all laboratories handled non-detect testosterone data as requested. Instructions were provided regarding how to handle non-detect sample results. Briefly, if a testosterone concentration below the detection limit was determined with a dilution factor of 10, then the sample was to be reanalyzed without dilution, i.e. with a dilution factor of 1. If a testosterone concentration below the detection limit was obtained with a dilution factor of 1, then the value was to be set to the detection limit of 0.2 ng/mL and treated as if it had been a measured value. Upon detecting testosterone samples below the detection limit, a couple of labs initially left the cell blank or marked the cell as below detection limit. This was corrected prior to their final submission of the spreadsheets.

Monitoring three of the five laboratories' conduct of the third replicate study indicated that there was sufficient differences in the elapsed time for key steps that the laboratories were requested to provide information about these steps. The elapsed time for these steps was summarized in the preceeding subsections.

4.0 DISCUSSION

The U.S. EPA implemented an Endocrine Disruptor Screening Program, and in this program, comprehensive toxicological and ecotoxicological screens and tests are being developed for identifying and characterizing the endocrine effects of chemicals with steroidogenic-altering activity. The sliced testis assay is one such screening test that has been proposed because it represents an assay that could be conducted at a minimal cost, quickly, and simply with standard laboratory equipment and basic laboratory training; the preparation is relatively stable as the fragments remain viable for several hours; the architecture of the organ is maintained; use of testis fragments allows a reduced number of animals to be used; the assay is relatively easy to standardize; and the assay has well-defined and multiple endpoints. In the present study, the intra- and interlaboratory variability when media, media-vehicle, positive control (AG at three concentrations), and cytotoxicant control (EDS at one concentration) groups were evaluated by measuring testosterone and LDH concentrations at 0 hours (baseline) and 0-4 hours (composite) after treatment.

The discussion is divided into sections that focus on the findings for the untreated groups (Media and Media-Vehicle) and treated groups (AG and EDS) as they pertain to the objectives of the work assignment.

4.1 Untreated Groups

Magnitude of Assay Response. The fragment response in the presence and absence of hCG is important information to obtain since the magnitude change between these two groups provides a measure of assay sensitivity. The magnitude change was calculated using the MV-control group data and dividing the testosterone concentration fragment response w/ hCG by the testosterone concentration w/o hCG to obtain a stimulated-to-unstimulated fold change. The values for this "S-to-U fold" parameter were calculated using the results from the individual laboratories by replicate (Table 3.5-5). The S-to-U fold values for Laboratories A through E were 7.4, 4.4, 11.7, 4.1, and 9.7, respectively. Using these values, the overall mean and SEM are 7.5 ±1.5. Based on mixed modeled analysis of variance fits to the logarithmic responses the S-to-U fold values were 7.3, 4.1, 11.1, 3.6, and 8.5, respectively (Table 3.4-1). The individual laboratories can be categorized into one of three groups - low, mid, and high. Labs B and D had relatively low S-to-U fold values, Labs A and E were in the mid range, and Lab C was in the high range. These findings raise two obvious questions: Are there reasons that might explain the differences in fragment responsiveness for the various laboratories? And, is the S-to-U fold parameter outcome sufficient to detect inhibitors? Responses to each of these questions are discussed in the following paragraphs.

Previous studies reported lower to similar values for testosterone production. In the present study, the (unweighted) average across laboratories of the unstimulated and stimulated hourly average testosterone concentrations were 0.16 and 1.1 ng/mg/hr. Powlin et al (1998) reported values of 0.3 and 0.9 ng/mg/hr. Laskey et al (1994) and Gray et al (1995) reported values for stimulated fragments and, after normalization to similar units, had results for stimulated fragments of approximately 0.5 and 0.4 ng/mg/hr. Direct comparison of the results is not possible since many of the experimental factors were not the same, e.g. fragment size, incubation time. Of the three investigators reported, Powlin et al., used an experimental design that most resembled the present study. Slightly improved values for the unstimulated fragment (lower) and stimulated fragment (higher) in the present study relative to the Powlin et al results are attributed to using optimized conditions in the present study.

Fragment responsiveness is affected by assay performance and laboratory performance. Assay performance factors were addressed in previous work assignments and involved assay optimization. The optimization study results were used to optimize age of the test system, media type and atmosphere, testis processing time, fragment size, incubation temperature and times, hCG concentration, sampling times, and aliquot volume, to name a few. The optimal assay conditions were specified and required to be used by all laboratories in the present study, thereby making it reasonable to examine laboratory performance as a possible explanation for a given measure of fragment responsiveness. Section 3.6, Laboratory Procedural Findings, identifies what were considered critical steps in the assay and provides information about the time taken to perform these critical steps. If indeed these steps are critical or have some bearing on the study outcome, then departure from the specified procedure and/or differences in performance among laboratories may be useful in helping to explain differences in the fragment responsiveness. The underlying factor believed most important to optimize fragment responsiveness was time - time to isolate, time to slice, time to process, time to sample collect, etc. since this time represents

how long a testis or fragment was not in conditions that favor maintaining viability, (e.g. optimal atmosphere, media, temperature). Therefore, laboratory performance outcomes were used to determine whether there was an association with the level of S-to-U fold parameter values. Labs B and D had the lowest S-to-U fold increase. Lab D was consistently slower in the time span taken to initiate testis fragmentation after collection, collect media samples, and conduct the assay from start to finish. Unfortunately, most of the information about time to execute a given step was not recorded by Lab B. These two laboratories used one technician to perform the vast majority of the steps of the assay once the testis was isolated, whereas the other laboratories used two or more technicians throughout most of the conduct of the assay. In contrast to the Labs B and D results, Lab C laboratory fragment responsiveness results were very consistent for all three replicates and indicated that a high S-to-U fold value could be achieved. A review of the Lab C laboratory performance results shows a consistent better-than or, at least, as-good-as elapsed time as any of the laboratories with the lowest elapsed time for completing the critical steps. Unfortunately, the WAL was not able to monitor Labs B or C and observe how the assay was being conducted so that comparisons could be made with the observations made at Labs A, D, and E. Thus, although a highly subjective evaluation, the laboratories with higher S-to-U fold values were associated with using less time for tissue processing and sample collection, thereby improving how long the fragments were in favorable conditions.

Although the magnitude of the S-to-U fold parameter was statistically significantly different for all laboratories, it was hoped that the response would be 10-fold or more. This outcome was achieved by one of the five laboratories and within approximately 7-fold or better for three of five laboratories. Laboratory performance, as described above, may explain the shortcomings of the other two laboratories. The target response was based on the notion that inhibitory effects of steroidogenic disruptors could be more readily detected if the difference between the unstimulated and stimulated testosterone production increased, assuming of course that the variability in the response was sufficiently low for an effect to be detected. However, even with the low S-to-U fold values reported by some of the laboratories, a statistically significant change was measured when the MV-control w/ hCG was compared to the AG groups. There was a concentration-dependent decrease in testosterone concentration that was statistically significant. For some laboratories, the S-to-U ratio was very low, but the inhibitory effect of AG was still detected (Table 3.4-1). Thus, while the sensitivity of the assay to detect changes at lower inhibitor concentrations or substances with less-inhibitory activity decreases when the S-to-U fold parameter decreases, inhibition of steroidogenesis, using AG, was still detected by the assay.

Variability of Assay. Measurements of assay reliability can be determined using the information obtained from the M-control w/ hCG and MV - control w/ and w/o hCG. The measurements include a) coefficients of variation (CV) across studies, b) ratio of between- to within-study standard error, and c) comparison of within-lab standard deviation to average within-lab standard deviation.

The CV within laboratories reflects the reproducibility of replicate results within a study for that laboratory. For the M-control w/ hCG and MV-control w/ and w/o hCG, the CV within laboratories ranged from 6 to 46 percent, 15 to 35 percent, and 12 to 52 percent, respectively

(Table 3.1-2). A target within laboratory CV value of 30 percent or less is preferred for biological assays. Two of five laboratories had within laboratory CV values of 30 percent or less for all three groups and one of five laboratories had within laboratory CV values of 30 percent or less for the M- and MV-control w/hCG groups but not the MV-control w/o hCG group, which had a value of 48 percent. The CV across laboratory values for the M-control w/ hCG and MV-control w/ and w/o hCG are 58, 50, and 35 percent (Table 3.1-2).

The ratio of the overall between standard deviation to the average standard error within laboratories reflects the relative contribution to total variation of the variability among study means as compared to the precision within studies. For the M-control w/ hCG and MV-control w/ and w/o hCG the ratios are 4.7, 3.8, and 1.7, respectively (Table 3.1-2). For two of the three ratios, the variability departed from unity by 4 to 5 fold. These results indicated that there is more heterogeneity in the average response across laboratories for the stimulated groups when compared to the unstimulated groups. The percentage of variability associated with heterogeneity among laboratory means for the M-control w/ hCG and MV-control w/ and w/o hCG were 95, 93, and 65 percent, respectively (calculated using the following equation: $\{[1-1/R^2] \times 100\}$, where R is the ratio listed above).

A comparison of the within-laboratory standard deviation to the average within-laboratory standard deviation measures the reproducibility of each laboratory relative to the average reproducibility of the laboratories. For the M-control w/ hCG and MV-control w/ and w/o hCG groups, the ratio of within standard deviation to the average standard deviation values for Lab A are 1.4, 1.0, and 0.5; for Lab B are 1.7, 1.4, and 1.9; for Lab C are 0.6, 1.2, and 0.4; for Lab D are 1.1, 1.0, and 1.6; for Lab E are 0.3, 0.5, and 0.7, respectively (values calculated using data in Table 3.1-2). Only one of five laboratories had ratios for all three groups below one, three of five laboratories had at least two of three ratios below or equal to one, and two of five laboratories had ratios for all three groups equal to or greater than one. These results indicate that the within laboratory variation is relatively consistent across laboratories. For each of these responses no laboratory had variability greater than twice the average.

The magnitude of the fragment response and variability assessment suggest that it is possible to perform the assay in a manner that produces satisfactory measures of response with acceptable variability, e.g. approximately 10 fold S-to-U paramter with CV percentages less than 30 percent. However, achieving these performance criteria occurred for a few rather than most of the laboratories, thereby indicating that factors that affect fragment response and variability were not controlled as planned or are not sufficiently understood to ensure that optimal assay performance is achieved.

4.2 Treated Groups - Positive and Cytotoxicant Controls

Two test chemicals were evaluated in this study - aminoglutethimide (AG), the positive control, and ethane dimethane sulfonate (EDS), the cytotoxicant control. The Lab A investigators, provided a brief description of the mode(s) of action of these two substances. AG, initially used as an anticonvulsant and identified as a goitrogen by inhibiting thyroxine synthesis

(Rallison et al., 1967), inhibits the first P450 enzyme in the intra-mitochondrial transformation of cholesterol to testosterone, i.e. P450 scc (side chain cleavage), which converts cholesterol to pregnenolone (Dexter et al., 1967). AG also inhibits 11β-hydroxylase in adrenal cells (Goldman, 1970); with this enzyme inhibited (which converts 11-deoxycortisol to cortisol), accumulating steroids are channeled into androgenic pathways. AG also inhibits aromatase, which converts testosterone or androstenedione to 17β–estradiol or estrone, respectively (Johnston, 1997; Yue and Brodie, 1997). AG may also interfere with hepatic metabolism of steroid hormones (Horky et al. 1969; 1971). EDS, an alkylating agent, has been shown to reduce 3β-HSD by 99 percent and serum testosterone in treated rats (Gray et al., 1995). Immature rat Leydig cells are less sensitive to EDS than adult Leydig cells (Kelce et al., 1991). Both mitogenic and steroidogenic activities in the testis are affected by EDS (Drummond et al., 1988). *In vivo*, the reduction in serum testosterone is paralleled by the loss of Leydig cells; by 9-10 days post dosing, neither serum testosterone nor Leydig cells are detectable in the male rat (Gray et al., 1995; Klinefelter et al., 1994). EDS may act by interfering with the cyclic AMP second messenger system, a signal transduction process that leads to steroidogenesis (Risbridger et al., 1989). Cyclic AMP stimulates the production of protein kinase A, which in turn, induces the synthesis of the cholesterol transport protein (steroid acute regulatory protein [StAR]), thereby transporting cholesterol into the Leydig cell mitochondrion from the outer membrane to the inner membrane. EDS-related interference with these processes inhibits gonadal steroid production and structural damage.

Magnitude of Assay Dose Response - Aminoglutethimide, the Positive Control. The present task evaluated aminoglutethimide as a positive control for the assay. In addition, although the sliced testis assay is being considered as a screening tool and not for determining possible mode(s) of action or demonstrating concentration-response relationships, the present study results indicated that the assay is capable of providing this information. AG produced a concentration-dependent decrease in testosterone concentration in stimulated fragments and this finding was consistent for all laboratories. Based on the percent decrease by replicate and laboratory for the M-V control +hCG and AG groups, testosterone concentrations decreased 45.0 ± 4.9, 84.4 ±2.0, and 96.2 ±0.6 percent relative to control for the 10, 100, and 1000 µM AG groups, respectively (Tables 3.5-6 through 3.5-8). Based on mixed modeled analysis of variance fits to the logarithmic response, the percentage inhibitory values were 45, 86 and 96 percent (Table 3.4-2). These values are similar to those above. The concentration-dependent results were unequivocal. It is interesting to note that even with the varying magnitude in fragment response and assay variability for the laboratories, all laboratories measured a decrease in testosterone concentration that was clearly attributed to AG. These results indicated that AG would be an effective positive control for the sliced testis assay.

Variability of the Assay - Aminoglutethimide. As described above, measurements of assay reliability can be determined using a) coefficients of variation (CV) across studies, b) ratio of between- to within-study standard error, and c) comparison of within-lab standard deviation to average within-lab standard deviation.

The CV values within laboratories for the 10, 100, and 1000 µM AG groups across laboratories ranged from 21 to 41 percent, 11 to 35 percent, and 13 to 45 percent, respectively

(Table 3.1-3). A target within laboratory CV value of 30 percent or less is preferred for biological assays. Only one of five laboratories had within laboratory CV values of 30 percent or less for all three groups and three of five laboratories had within laboratory CV values of 30 percent or less for two of three AG concentrations. These ranges are similar to what was obtained for the M- and MV-control groups. The CV across laboratory values for the 10, 100, and 1000 µM AG groups are 65, 63, and 32 percent (Table 3.1-3).

The ratios of the overall between standard deviation to the average standard error within laboratories are 4.3, 4.1, and 2.1 for the 10, 100, and 1000 µM AG groups, respectively (Table 3.1-3). For two of the three ratios, the interlaboratory variability departed from unity by approximately 4 fold. These results indicated that there is slightly more heterogeneity in the average response across laboratories for the low and mid concentration groups when compared to the high concentration group. The percentage of variability associated with heterogeneity among laboratory means for the 10, 100, and 1000 µM AG groups were 84, 83, and 31 percent, respectively (calculated using the following equation: $\{[1- 1/R^2] \times 100\}$, where R is the ratio listed above).

For the 10, 100, and 1000 µM AG groups a comparison of the within-laboratory standard deviation to the average within-laboratory standard deviation for Lab A are 1.1, 0.9, and 1.4; for Lab B are 1.0, 0.5, and 0.8; for Lab C are 1.6, 2.6, and 1.4; for Lab D are 0.6, 0.6 and 0.6; and for Lab E are 0.7, 0.5, and 1.0, respectively (values calculated using data in Table 3.1-3). Three of five laboratories had ratios for all three treatment groups equal to and below one and one of five laboratories had ratios for all three groups greater than one. These results indicate that the within laboratory variation is relatively consistent across laboratories.

These results indicated that the magnitude of the fragment response was sufficient and variability low enough for the inhibitory effect of AG to be measured in a concentration-dependent fashion across all laboratories. Based on these results, AG is considered an acceptable selection as a positive control for the assay.

Magnitude of Assay Cytotoxicant Response - EDS, the Cytotoxicant Control Control. EDS was tested at a single concentration, 1000 µM, and while it significantly decreased the testosterone concentration there was no concomitant increase in LDH concentration. Based on the percent inhibition by replicate and laboratory for the M-V control +hCG and EDS groups, the relative percentage decrease in testosterone concentration was 76.6 ±2.0 percent (Table 3.5-9). Laskey et al (1994) and Gray et al (1995) reported that EDS has a concentration- and time-dependent effect on the testosterone concentration. However, selection of a cytotoxicant control necessitates that the candidate cytotoxicant produce measurable changes in an endpoint that can be used to assess tissue and/or cellular viability, which for this study was LDH. Laboratory LDH values following EDS treatment were similar to the M- and MV-control group values (Tables 3.2-2 and 3.2-3). Thus, it was not possible from the results of the present study to demonstrate that a cytotoxicant effect was produced, thereby making selection of EDS as the cytotoxicant control tentative pending further study.

Variability of the Assay - EDS. The within laboratory CV values for 1000 µM EDS ranged from 4 to 27 percent and the across laboratory CV value was 18 percent (Table 3.2-3). A target CV value of 30 percent or less is preferred for biological assays. These ranges are better than what was obtained for the M- and MV-control and AG groups. The ratio of overall between standard deviation to the average standard error within laboratories is 2.1 (Table 3.2-3). The within-laboratory standard deviation relative to the average within-laboratory standard deviation was 1.0, 1.5, 0.3, 1.4, and 0.8 for Labs A through E, respectively (values calculated using data in Table 3.2-3). Three of five laboratories had ratios equal to or below one.

These results indicated that the magnitude of the fragment response was sufficient and variability was low enough for the testosterone inhibitory effect of EDS to be measured but the absence of being able to measure cytotoxicity precludes accepting EDS as the cytotoxicant control without further investigation.

4.3 Interlaboratory Discussion

The interlaboratory statistical analysis combined summary values developed by each of the laboratories in their individual intralaboratory statistical analyses. It assessed the extent of variation among results reported by the laboratories, overall consensus estimates across laboratories, and the presence of any outlying laboratories. The interlaboratory analyses were based on (natural) logarithmically transformed baseline adjusted composite concentration comparisons between each of the AG inhibition groups, cytotoxicant inhibition group, and alternate control groups and the stimulated media vehicle control group.

The first order conclusion is that each of the five laboratories had qualitatively the same results. Each laboratory observed successively decreasing testosterone concentration inhibition as the concentration of AG inhibitor increased. The response trend in each laboratory was well approximated by a decreasing straight line on the logarithmic testosterone scale relative to the logarithm of AG concentration. Furthermore each laboratory observed statistically significant inhibition in testosterone concentration in the unstimulated control group and the cytotoxicant control group but not in the stimulated media control group.

The laboratories were also in agreement with respect to the LDH concentration results. There were no statistically significant changes relative to the stimulated media-vehicle control group except for several isolated comparisons which appear to be random.

5.0 CONCLUSIONS

The sliced testis assay was conducted by five laboratories to determine testosterone and LDH concentration baseline levels and effects of treatment with positive (AG) and cytotoxicant (EDS) controls. Measures of variability were determined for control and treatment groups within and across laboratories. AG, at concentrations ranging from 10 to 1000 µM, produced a statistically significant concentration-dependent decrease in testosterone concentration across all

laboratories and within all laboratories. AG treatment had no effect on LDH. EDS, at a concentration of 1000 µM, significantly decreased testosterone concentration, but not LDH, across all laboratories and within all laboratories. AG would be an effective positive control test substance for the assay; however, EDS was equivocal as a cytotoxicant based on LDH as the endpoint for assessing cytoxicity. Interlaboratory statistical analysis showed that each of the five laboratories had qualitatively the same results for testosterone and LDH by group.

6.0 REFERENCES

DerSimonian, R. and Laird, N. *Meta-analysis in clinical trials*, Controlled Clinical Trials, 1986, Vol 7, p. 177-188.

Dexter, R. N., Fishman, L. M., Ney, R. L., *et al.* (1967). Inhibition of adrenal corticosteroid synthesis by aminoglutethimide: Studies of the mechanism of action. *J. Clin. Endocr.* **27**, 473-480.

Drummond, A. E., Bisbridger, G. P., O'Leary, P. C., and de Kretser, D. M. (1988). Alterations in mitogenic and steroidogenic activities in rat testicular interstitial fluid after administration of ethane dimethane sulphonate. J. Reprod. Fertil. 83:141-147.

Goldman, A. S. (1970). Experimental congenital lipoid adrenal hyperplasia: Prevention of anatomic defects produced by aminoglutethimide. *Endocrinology* **87**, 889-893.

Gray, L. E., Klinefelter, G., Kelce, W., Laskey, J., Ostby, J., and Ewing, L. (1995). Hamster Leydig cells are less sensitive to ethane dimethanesulfonate when compared to rat Leydig cells both *in vivo* and *in vitro*. *Toxicol. Appl. Pharmacol.* **130**, 248-256.

Hartung, J. and Makambi, K.H. *Simple non-iterative t-distribution based tests for meta-analysis*. South African Statistical Journal, 2001, Vol. 35, p. 1-17.

Horky, K., Kuchel, O., Gregorva, I., *et al.* (1969). Qualitative alterations in urinary 17-ketosteroid excretion during aminoglutethimide administration. *J. Clin Endocrin.* **29**, 297-299.

Horky, K., Kuchel, O., Starka, L., *et al.* (1971). Effect of aminoglutethimide on extraglandular metabolism of exogenous testosterone. *Metabolism* **20**, 331-336.

Johnston, J. O. (1997). "Aromatase Inhibitors." In: *Biochemistry and Function of Sterols*. Eds. Parish, E. J. and Nes, W. C. CRC Press, pp 23-53.

Kelce, W. R., Zirkin, B. R., and Ewing, L. L. (1991). Immature rat Leydig cells are intrinsically less sensitive than adult Leydig cells to ethane dimethanesulfonate. *Toxicol. Appl. Pharmacol.* **111**, 189-200.

Klinefelter, G. R., Laskey, J. W., Kelce, W. R., Ferrell, J., Roberts, N. L., Suarez, J. D., and Slott, V. (1994). Chloroethylmethanesulfonate-induced effects on the epididymis seem unrelated to altered Leydig cell function. *Biol. Reprod.* **51**, 82-91.

Laskey, J. W., Klinefelter, G. R., Kelce, W. R., and Ewing, L. L. (1994). Effects of ethane dimethanesulfonate (EDS) on adult and immature rabbit Leydig cells: comparison with EDS-treated rat Leydig cells. *Biol. Reprod.* **50**, 1151-1160.

Powlin, S. S., Cook, J.C., Novak, S., and O'Connor, J.C. (1998). Ex vivo and in vitro testis and ovary explants: Utility for identifying steroid biosynthesis inhibitors and comparison to a Tier I screening battery. Toxicological Sciences, 46:61-74.

Rallison, M. L., Kumagai, L. F., and Tyler, F. H. (1967). Goitrous hypothyroidism induced by amino-glutethimide, anti-convulsant drug. *J. Clin. Endocr.* **27**, 265-272.

Risbridger, G., Kerr, J., and De Kretser, D. (1989). Differential effects of the destruction of Leydig cells by administration of ethane dimethane sulphonate to postnatal rats. *Biol. Reprod.* **40**, 801-809.

Yue, W. and Brodie, A. M. (1997). Mechanisms of the actions of aromatase inhibitors 4-hydroxyandrostenedione, fadrozole, and aminoglutethimide on aromatase in JEG-3 cell culture. *J. Steroid Biochem. Mol. Biol.,* **63**:317-328.