SUMMARY OF SAFETY AND EFFECTIVENESS DATA

I. General Information

Device Generic Name: Salivary biochemical marker for risk assessment of spontaneous preterm labor and preterm delivery

Device Trade Name: Salest™

Applicant's Name and Address:

Biex, Inc.
6693 Sierra Lane
Dublin, CA 94568

Premarket Approval Application (PMA) Number: P970032

Date of Panel Recommendation: December 10, 1997

Date of Notice of Approval to Applicant APR 2 9 1998

II. Indications for Use

The Salest™ test is intended to detect and measure by enzyme-linked immunoabsorbant assay (ELISA) technology the level of salivary estriol in pregnant women. The device is indicated for use as an aid in identifying risk of spontaneous preterm labor and delivery in singleton pregnancies. The device can be used every 1 to 2 weeks from gestational ages 22 to 36 weeks. The test should be used as a component of the clinician's assessment of risk for preterm labor and delivery.

III. Device Description

The device is composed of two main elements: the Salivary Estriol Enzyme-Linked Immunoabsorbant Assay (ELISA) Test Kit (Salest™) and the Saliva Collection Kit.

Collection Kit

The collection kit is designed to enable the appropriate sample to reach the testing laboratory in adequate quantity without degradation.

The collection kits are available in two configurations: a 4-pack collection kit (containing 4 collection units) and single collection kit. Each collection unit contains a funnel, capped collection tube, plunger with preservatives and surfactants, shipping envelope, plastic shipping case, directions for use, patient education booklet, and reminder kit.

The collection tube consists of a 13 mm X 100 mm test tube with a screw-cap closure. The tube has a label that includes an indicator band that serves to provide visual confirmation that sufficient volume, 1 mL, has been provided. The label also contains a bar code that identifies the tube. The collection tube also consists of a filter plunger device that slides into the test tube after the sample has been provided. The plunger serves the dual purpose of filtering and introducing into the sample antimicrobial agents to prevent overgrowth of microorganisms found in the oral cavity. After sealing the collection tube with the plunger in place, the filter creates an inner tube sealed on one end by the filter and on the opposite end by the screw cap.

The mailer is designed to protect the sample tube from danger of crushing during transport to the laboratory. In addition, the mailer contains an absorbent material that will retain the saliva sample in the event the tube leaks for any reason. The tube and absorbent material are contained within two separate leak-proof barriers to meet the requirements of the International Air Transportation Authority (IATA) for the shipment of biological specimens.

The Biex Salest™ Test Kit

The Biex Salest™ test is a competitive polyclonal antibody based microplate enzyme immunoassay. Estriol in the sample competes with an estriol enzyme conjugate for limited antibody binding sites on the solid phase microwell. Polyclonal rabbit anti-estriol is coated onto the surface of microplate wells. Prior to adding calibrators or patient saliva samples a quantity of Pretreatment Buffer is added to each well to prepare the well for the raw saliva sample. After pretreatment of the microplate wells calibrators containing increasing amounts of estriol and patient saliva samples are added to appropriate antibody coated wells of the microplate. Horseradish peroxidase (HRP) conjugated to estriol is added to each well and mixed. The plate is then incubated for one hour. The estriol in the calibrators and saliva samples compete with the enzyme conjugate for binding to the anti-estriol antibody coated onto the microplate wells. The wells are then washed with a specially formulated wash buffer to remove unbound material. After washing HRP substrate is added to each well and incubated for 20 minutes. The reaction is stopped with dilute hydrochloric acid. The color development is then read on a microplate reader. The concentration of estriol in the samples is determined using a calibration curve.

The Salest™ kit includes Anti-Estriol Microwell Strips, Enzyme Conjugate, substrate, stop solution, Estriol Standards (0.0, 1.5, 3.0, 5.0, 10.0 ng/mL), pretreatment buffer, and wash buffer concentrate.

Test Procedure - Specimen Acquisition

8

A saliva sample is provided every one or two weeks, by a pregnant woman beginning at 22 to 24 weeks gestation, either at her physician's office or clinic or at home through week 36 of gestation. Samples are mailed to the testing laboratory where samples are tested with the Salest™ test. Test results are reported within 24 hours of their receipt in a variety of formats (e-mail, fax, phone or mail) depending on physician preference.

CONTRAINDICATIONS:

There are no known contraindications for the Salest™ test.

WARNINGS AND PRECAUTIONS

Warnings and precautions for use of the device are stated in the attached product labeling. (Attachment A)

III. Alternative Practices and Procedures

Risk for preterm labor is currently assessed on the basis of a patient's medical history, lifestyle, behavior, and demographic profile. Several scoring systems have been developed to identify women at increased risk. The most commonly used are the Creasy-Papiernik system and guidelines provided by the American College of Gynecologists (ACOG). [1-4] In addition, practitioners frequently assess whether a major risk factor for preterm labor is present. [5] Such assessments are usually performed in an interview style at the first prenatal visit and again later in gestation. Table 1 lists some of the typical risk factors associated with preterm labor.

Table 1 Risk Factors Associated with Preterm Labor
ACOG Technical Bulletin, October 1989

Prior preterm delivery	Multiple gestation
3 or more first-trimester abortions	Previous second trimester abortion
Cervical incompetence	Abdominal surgery during current pregnancy
Uterine or cervical abnormalities	Premature placenta separation
Placenta previa	Hydramnios
Fetal abnormality	Second-trimester bleeding
Serious maternal infection	Pregnancy weight < 100 lbs.
Cervical effacement or dilation (> 50% or > 1 cm.)	Single parent
No prenatal care	

In addition to the above, current management protocols might also incorporate new etiology-specific risk assessments tools such as fetal fibronectin or transvaginal cervical ultrasound that recognize the physiological changes which signal impending preterm labor. [5]

V. Marketing History

This device does not currently have any marketing history either in the U. S. or outside the U. S.

VI. Adverse Effects of Device on Health

A false positive result could result in the patient being monitored more rigorously for onset of preterm labor when such monitoring may not be needed. In some cases the patient might be advised to take bed rest (life-style/employment implications) unnecessarily. When evaluated with other patient risk factors a false-positive result could be a determining factor in making a decision to administer tocolytic or steroid drugs to prevent the onset of labor. Alternatively, a false negative test result could result in a decision not to implement certain actions (as above) to prevent preterm labor.

VII. Summary of Studies

Non-clinical Studies

Laboratory Studies

Sensitivity

The sensitivity of the test, defined as the lowest concentration of estriol that can be distinguished from the 0 ng/mL Calibrator, is 0.19 ng/mL.

Precision

A precision study was conducted to determine the within-run, between-run, and total precision for the kit calibrator and control optical absorbances, and the concentrations of the controls.

The assay was conducted for 14 days using a fresh kit each day from each of three lots. Standard curves were generated using duplicates of each standard and low-, mid- and high-level estriol controls, and five replicates were run for each test. The total precision for control concentrations over all runs was found to be 11.3% or less. The within-run, between-run and lot-to-lot variability determined for the Salest™ ELISA is acceptable for a test of this type.

Recovery

This study was designed to determine the accuracy of recovery of the Salest™ test when a known amount of estriol was added to saliva. Two separate analyses were performed. The first was a linear regression analysis of the standard curves of spiked samples. The

second used a single spiked level in multiple samples and a t-test on multiple replicates to determine if there was significant difference between the observed and expected values.

Based on the criterion that a slope of 1. 0 ± 0.1 is acceptable, all three lots were acceptable in terms of spike recovery.

Table 2 Spike Recovery
Linear Regression of Standard Curves

Lot	Slope	Y-Intercept	Correlation Coefficient
1	1.1	-0.070	1.000
2	1.1	-0.167	0.982
3	1.0	-0.086	0.990

The results of the second study of 40 patient samples demonstrated that there was 95% (1.9/2.0) recovery.

Linearity

A Saliva Pool containing a known amount of estriol was serially diluted with Calibrator diluent. The resulting linear regression equation was $y = 0.9909x + 0.1389$, $R^2 = 0.9969$. The observed and expected results demonstrated acceptable linearity after dilution.

Specificity

The Biex Salest™ test was specific for estriol. There was cross-reactivity with 16 epiestriol, estriol-3-sulfate, estriol-3-glucuronide, and estriol-16-glucoronide. There was little or no cross-reactivity with other hormones tested. Table 3 contains the list of hormones tested.

Table 3 Cross Reactivity

COMPOUND	% CROSS REACTIVITY
estriol-3-glucuronide	52
estriol-3-sulfate	60
estriol-16-glucuronide	6
estrone-3-glucuronide	< 1
estrone-3-sulfate	< 1
estradiol	< 1
17-a-estradiol	< 1
17-epiestriol	< 1
16-epiestriol	11
cortisone	< 1
11-deoxycortisol	< 1
5-a-androstenedione	< 1
digoxigenin	< 1
digoxin	< 1
progesterone	< 1
pregnanediol glucuronide	< 1
testosterone	< 1
estriol	100

Preclinical Studies

A series of preclinical tests were conducted to ensure that the performance and effectiveness of the Salest™ test were not compromised by the potential variations that might occur during normal use of the test.

Biex evaluated short-term (< 20 minutes) sequential sample collection, possible diurnal variations in estriol levels, and sample handling and stability. Further, the potential for assay interference by food and drink, smoking, tooth brushing, and bleeding gums resulting from tooth brushing were also conducted.

Diurnal Variation

One analysis indicated that the variation in the hourly values from 9:00 a.m. to 8:00 p.m., inclusive, was not statistically significantly different from the variation in the values collected 15 minutes apart (where diurnal variation was not a factor). This finding was consistent with the absence of diurnal variation expected between the hours of 9:00 a.m. and 8:00 p.m.

A second analysis further indicated that specimens collected intra-patient between 8:00 a.m. and 10:00 p.m. were not statistically significantly different as a result of diurnal

variation. As an additional safety factor Biex recommended in package labeling that saliva specimens be taken between 9:00 a.m. to 8:00 p.m.

Effects of Food and Drink

Based on some of the results, it would appear that ingestion of food or drink lowers somewhat salivary E3 levels from pre-consumption baseline levels. Although food and drink appear to have a minimal impact on salivary levels, it is recommended in the package labeling that patients wait at least one hour following consumption of any food and drink before taking a saliva specimen.

Effects of Smoking

Because of patient to patient differences observed in the ratio of post-smoking to pre-smoking estriol levels, it is recommended that the user avoid smoking at least one hour before taking a saliva sample.

Effect of Tooth Brushing

The analysis revealed no significant differences between the baseline value and the sample value at any time point except the one occurring at 90 minutes. This single finding was inconsistent with the others and therefore appeared to be a random outlier and clinically irrelevant. The labeling instructions conservatively require waiting an hour after tooth brushing before taking a sample.

Effect of Tooth Brushing and Mouth Washing

The results of this study indicated that patients should not submit a specimen collected within 10 minutes of using mouthwash. The labeling instructs patients to wait at least one hour following tooth brushing and mouth washing before taking a saliva specimen.

Effects of Bleeding Gums Due to Tooth Brushing

The results of the analysis revealed a significant difference ($p < 0.05$) from baseline at 10 and 30 minutes post-brushing in the bleeding gums group. For the group of patients with non-bleeding gums, there was a difference from baseline at 90 minutes post-brushing, but this was equivocal. The rise at 10 minutes, while not significant, was similar to the rise observed in the bleeding gums group shortly after brushing. Therefore, waiting an hour after tooth brushing is advisable and the device is so labeled.

Sequential Sampling

Three or more sequential samples within 10 minutes should be avoided because the evidence demonstrated a trend of significantly lower salivary estriol values with repeated collections. Product labeling reflects this aspect.

Preservative, Filtering and Freezing Effects

There was no evidence of any difference associated with specimen handling methods.

Short-term Storage

There was no difference in salivary estriol level at 30 minutes, 1, 2, or 24 hours of sample storage. These data supported both the current testing protocol where samples are mailed to a laboratory via the U.S. Postal System, and immediate testing such as at a point of care location.

The 3-hour and 4-hour values showed a statistically significant difference from the initial value. Samples tested during this period may result in artificially low estriol values. Although this decrease may have been due to some function of the saliva matrix itself, this limitation has been included in the labeling.

Feasibility Study Retest

The purpose of this test was to evaluate the effect of modifying the sample by introducing a pretreatment buffer to the test. Specifically, the test examined the possible effect on test cut-off. The performance of the kit subsequent to the buffer pre-treatment procedure was revalidated against original test results. This was done by reanalyzing a sub-set of samples (from 87 patients) of the total original samples (from approximately 250 patients) that were collected during the feasibility clinical trial.

Based on ratio analysis of paired values and regression analysis, the two methods (original and modified) were comparable. Additionally, the clinical sensitivity and specificity for the two methods were very similar. Therefore, it can be concluded that there was no clear significant difference between the two methods.

Clinical Studies

The pivotal study was a triple-blinded, multi-site, longitudinal, prospective evaluation of salivary estriol (E3) as a risk assessment marker for spontaneous preterm labor in women with singleton pregnancies. The study was triple blinded in that the E3 levels were not made available to the patient, physician, or sponsor conducting the laboratory analysis until completion of the study. Further, no clinical information was available to the laboratory personnel. The study was conducted in both low- and high-risk women as classified by the Creasy scoring system. For purposes of the analysis of effectiveness, the primary clinical endpoint was the incidence of spontaneous preterm labor and delivery in women with a salivary E3 level ≥ 2.1 ng/mL as compared to the incidence in women with an E3 < 2.1 ng/mL with a term delivery. Preterm labor was defined as spontaneous preterm labor resulting in preterm birth within 72 hours prior to 37 weeks gestation. The

cut-off of 2.1 ng/mL was established in the feasibility study, prior to the start of the pivotal study.

Summary of the Investigation

Patient Group

The study group included all women seeking prenatal care at participating study sites. A goal of the study was to enroll diverse ethnic backgrounds, ages, marital status, education levels, and economic status. Pregnant women receiving prenatal care meeting all the inclusion criteria and none of the following exclusion criteria were candidates for enrollment in this study.

Inclusion Criteria

- Eighteen years of age or greater;
- Creasy score < 10 for low and ≥ 10 for high-risk patients;
- Between 21 and < 25 weeks gestational age based on clinical information and evaluation of first ultrasound as described in gestational age determination protocol;
- No symptoms of preterm labor (PTL) upon study enrollment (≥ 4 uterine contractions/hour and cervix dilated ≥ 2 cm or a change of 1 cm or > 50% effaced). Uterine contractions or lack of uterine contractions at the baseline visit as determined by the patient and verified by the examiner with a cervical examination to distinguish between PTL and uterine contractions without PTL. A monitor was used upon request of the physician.
- No tocolytic therapy in the current pregnancy prior to study enrollment;
- Singleton pregnancy; and
- No known maternal illicit drug use or > 4 drinks/week of alcohol at the time of enrollment.

Exclusion Criteria

- Patient unwilling or unable to sign informed consent;
- Pregnancy complicated by known placenta previa or abruption;
- Cervical incompetence requiring cerclage;
- Major maternal medical complications, such as chronic hypertension, insulin dependent diabetes mellitus, or serious cardiac, renal, immunologic, hepatic, mental, digestive, endocrinologic, respiratory disease;
- Known fetal anomalies or intrauterine growth retardation;
- Known maternal drug or alcohol abuse;
- Presence of ruptured fetal membranes prior to enrollment;
- Pregnancy induced hypertension (PIH) upon study enrollment;
- Any medication known to affect hormone levels, i.e. Dilantin, steroid preparations, and Haldol;
- Signs/symptoms of preterm labor; and
- Gingivitis - bleeding gums that would interfere with sample collection.

Patient Enrollment and Baseline Data

Patients were enrolled at 21 to < 25 weeks gestation, and saliva samples were scheduled for collection weekly (every 2 weeks at clinic visits, in-between at home with samples sent in mailers to clinic) until delivery. Clinic visits were scheduled every 2 weeks until delivery.

The nurse coordinator evaluated each woman enrolling for prenatal care prior to 25 weeks gestational age using the gestational age determination criteria. Gestational age was verified as per the protocol and discussed below. The chart of each woman was reviewed for eligibility and to ensure that no exclusion criteria existed. Women had a Creasy Score determination using the Creasy Risk Assessment system described in Table 4. The woman was assigned a low-risk determination if her total score was < 10 and a high-risk if her total score was ≥ 10. If the woman met the other eligibility requirements, she was contacted and informed consent requested.

Table 4
Creasy Risk Assessment System
(Total Score: <10 Low-risk, ≥ 10 High-risk)

Pts	Socioeconomic Status	Past History	Daily Habits	Current Pregnancy
1	Low status; 2 children at home	1 abortion ≤ 1 yr last delivery	Works outside home	Unusual fatigue
2	Single; age: <20, >40 yrs	2 abortions	>10 cigarettes a day	Wt. Gain <12 lbs; hypertension; albuminuria; bacteriuria
3	Very low status; height: < 5 ft, weight: < 100 lbs.	3 abortions	Heavy work-, Long tiring trip	Breech; wt loss > 4 lbs; head engaged; febrile illness
4	Age: < 18	Pyelonephritis		Bleeding after 12 wks; effacement; dilation; uterine irritability
5				Placenta Previa; hydramnios
10		Uterine anomaly; 2nd trimester abortion; DES exposure; premature delivery; repeat 2nd trimester abortion		Twins; abdominal surgery

After the informed consent was signed, each woman was interviewed and her chart reviewed to collect demographic data, medical history, previous pregnancy history, and current data regarding pregnancy status such as signs and symptoms of vaginal infection or urinary tract infection. These data were recorded on the Case Report Forms. The data fields collected were:

Demographic Data

- Age
- Marital status
- Ethnic background - European, Asian, Hispanic, Native American Indian, African American, Other
- Years of education completed

Prior Pregnancy History

- Number of total pregnancies (gravida)
- Number of preterm and term pregnancies
- Number of spontaneous/therapeutic abortions
- Gestational age at completion of each prior pregnancy
- Number of living children

Behavioral Data

- Smoking
- All current medications including prescription and non-prescription medications.

Medical History

The medical history was reviewed for chronic hypertension, insulin dependent diabetes mellitus, HIV positive, serious cardiac, renal, hepatic, mental, neurological, digestive, endocrinologic, respiratory, or immunologic disease, urinary tract infections or other vaginal infections, including gonorrhea, chlamydia, trichomonas, bacterial vaginosis, group B streptoccus, and candida. The presence of these was recorded. History of abdominal or uterine surgery was also recorded. Any other medical condition for which the patient was receiving therapy was also recorded along with the treatment.

Current Pregnancy History

- Reported pre-pregnancy weight
- Weight at study enrollment
- Height
- LMP date
- Expected date of consignment (EDC)
- EDC by ultrasound

Baseline Examination

Upon completion of the history evaluation all women had a baseline physical examination. The baseline examination included signs and symptoms of PTL (i.e., presence or absence of abdominal pain, backache, flu-like symptoms, pelvic pressure, cramping, vaginal bleeding, diarrhea, and change in vaginal discharge). Evaluation of

any contractions or leaking of amniotic fluid was done to rule out PTL or rupture of membranes. A digital vaginal examination was performed to assess lower uterine segment, and the consistency, position, and dilation of the cervix. Cervical length was recorded as well as station of the presenting part.

Cervical Examination Procedure

The length of the cervix was estimated by digital examination from the lower uterine segment to the tip of the cervix in centimeters (to the nearest 0.5 cm) along both sides of the cervix at 3 and 9 o'clock. The average of the two sides was considered the estimated length. Dilatation was determined by digital examination of the external os using the forefinger. The transverse diameter of the external os at approximately 0.5 cm from the end of the cervix was estimated in cm to the nearest 0.5 cm. A paper ruler was available to estimate the cervical length and dilatation.

The following definitions were used during this assessment:

- Position: The cervical os was classified as anterior, mid position or posterior.
- Consistency: The cervix was classified as being soft, medium or firm.
- Lower uterine segment: The lower uterine segment was classified as normal or bulging.
- Station: Station was classified as high out of the pelvis, ballottable being able to feel the presenting part, or engaged when the presenting part is lodged in the pelvis.

Saliva Collection

At the enrollment visit (visit #1) each woman was instructed in the proper collection technique. The first saliva sample (total, unstimulated) was then collected. Subsequent collections were unsupervised whether at clinic or home.

Each study subject was then provided written instructions for performing subsequent saliva collections. (The patient was instructed to rinse her mouth with water, wait 10 full minutes then drool 2 ml of saliva into one of the collection tubes provided.) The patient was requested to collect a saliva sample between 9 a.m. and 8 p.m. and to avoid food, drink, smoking, tooth brushing, mouth washing, chewing gum or paraffin for at least one hour prior to sample collection. As previously stated saliva samples were scheduled for collection weekly (every 2 weeks at clinic visits, in-between at-home with samples sent in mailers to clinic) until delivery.

A checklist regarding sample collection was provided to each patient. The completed list was returned with the sample in the mailer. Provisions were made for collecting samples in the hospital or non-participating hospitals by supplying the patient upon enrollment with an extra saliva collection set and sample mailers.

Clinic Visits Schedule

A baseline visit occurred at 21 to < 25 weeks gestation followed by scheduled clinic visits at 25, 27, 29, 31, 33, 35, 37, 39, and 41 weeks gestation. Patients rendered saliva specimens up to when they delivered. At each visit the woman was carefully questioned as to any signs or symptoms of PTL. Cervical examinations were performed at the baseline visit and every four weeks at 27, 31, and 35 weeks gestation or more frequently as clinically indicated by the physicians caring for the patients. Ten clinic visits were scheduled; 4 of these visits required digital pelvic examinations.

At every visit the following occurred:

- The study coordinator ensured that home saliva samples had been collected.
- The woman was asked specifically about signs and symptoms (or absence) of PTL, specifically uterine contractions and rupture of membranes (ROM) were recorded.
- Any change in medication such as antibiotics and tocolytics was recorded.
- Any hospitalization, ER visits or screening visits were recorded.
- Saliva was collected.

Further, digital pelvic examinations were performed according to a schedule or when clinically indicated.

Patients were asked to continue providing saliva samples even if hospitalized. Symptoms of PTL prior to hospitalization were recorded. The presence, characterization (irregular or regular) and frequency of uterine contractions were recorded. Other recorded data included date of admission, diagnosis at admission, date of discharge, evidence of rupture of membranes, clinical diagnosis for PTL with intent to treat, any change in tocolytic treatment, cervical exam (if performed) results, medication changes, and date and time of saliva collection. If the woman was hospitalized and treated for PTL, a saliva sample was collected prior to initiation of tocolytic treatment or other intervention. If the woman was hospitalized or delivered outside of the study center, attempts were made to obtain her medical records and/or delivery information. A provision also was made for collecting saliva samples in the non-participating hospitals.

To ensure appropriate and consistent patient-to-patient measurements, key clinical end points including preterm delivery (PTD), PTL, premature rupture of membranes prior to 37 weeks (pPROM) and gestational age determinations were defined and used by all clinicians. In addition, sample collection methods and timing were clearly specified. The following clinical definitions were used during the study:

- PTD - A delivery prior to 37 weeks gestation.
- PTL - Labor < 37 weeks gestation with regular contractions (with or without cervical change) and a decision to tocolytic treat; or regular contractions (4 or more an hour) and cervical change of at least 2 cm or ≥ 50% effaced.

- Indication of pPROM - Rupture of membranes at less than 37 weeks without regular uterine contractions.
- The primary endpoint PTL was defined as spontaneous PTL (contractions starting with intact membranes) at < 37 weeks of gestation resulting in a PTD within 72 hours.

Criteria used to determine gestational age for enrolled patients were as follows:

- The last menstrual period (LMP) was based on the patient's recollection.
- If for any reason the LMP was judged equivocal, the ultrasound measurements obtained during a previous examination were used to determine the gestational age using the standard method of ultrasound gestational age determination at each institution.
- If the patient was sure of the date of her LMP it was used if the ultrasound agreed within < 14 days.
- If the ultrasound determined gestational age was different by more than 14 days from the patient determined LMP the ultrasound result was used to determine the gestational age.

The original gestational age entered on the enrollment document following the above guidelines was not to be changed once a determination was made.

Delivery Data

All enrolled subjects had medical records reviewed for delivery information such as date and time of delivery, type of delivery, occurrence of PTL, use of Betamethasone, use of tocolytics, gestational age at delivery and infant birth-weight gender and apgar scores. Other information recorded included time and date of rupture of membranes, and number of Intermediate Nursery and Neonatal Intensive Care Unit (NICU) days.

Sample Handling

Sample tubes were numbered with patient identification codes discussed below. Lab personnel were masked to these codes. Home samples (obtained between clinic visits) were collected by the patient and sent to the principal investigator soon after the collection in mailers. Home, clinic and hospital collected saliva samples were stored at the clinic at -20^0 C and, after the women delivered, were sent frozen to Biex, Inc. for batch assaying.

Blinding of Samples and Clinical and Laboratory Staff

Prior to shipment each sample was marked at the sites with a randomly generated computer number for masking purposes. The corresponding "code" number was placed on the Case Report Form (CRF) and sent to the data processing office and to the Biex testing facility. An off-site contractor processed the data following completion of the trial and data lock-up in February, 1997. Each site received pre-generated sequential patient identification/tracking numbers and each enrolled patient was assigned a number that was tracked on an enrollment form contained in the critical document notebook. This

enrollment number was also placed on all the CRFs and saliva collection tubes. Each patient number used at the site was tracked using Paradox computer software.

Data Quality Assurance

A device accountability log and a log of any patient complaints or problems was maintained during the entire course of this study. Data quality was maintained by ensuring that source documents and corresponding data were thoroughly cross-checked. Each site had a study nurse and a sponsor Clinical Research Associate monitoring the site on a monthly basis. In addition, an outside agency conducted Good Clinical Practices (GCP) audits of the clinical trial procedures, as well as record keeping and document tracking practices to further ensure accuracy and completeness of all study reports and documentation.

Removal of Patients from the Study

Each patient enrolled in the trial was evaluated for compliance to the protocol at the time of either completing or withdrawing from the study. If the patient withdrew from the trial and was considered *non-evaluable,* the patient was not considered in the analysis. If the patient missed a single sample collection but otherwise complied with the protocol she was considered in the analysis. Following all reasonable attempts to retrieve missing data, if the data could not be found or reconciled, the data field remained blank and no method of imputation was used. Throughout the summary reports of the data, the incidence of blank or empty data fields were reported and listed under the category of not reported.

Women were withdrawn from the study if it was determined they were unknowingly enrolled subject to an exclusion criterion. As an example, women who were enrolled and subsequently diagnosed with twins were withdrawn from the study. Patients were also withdrawn upon their request, if a saliva sample was not rendered for 3 consecutive weeks from the time of enrollment to the time of delivery, or if the patient missed 2 consecutive clinic visits. Women who missed one clinic visit or who missed cervical exams were not removed from the study but those not following the procedure for saliva collection were removed from the study. (Note: A single saliva sample not properly obtained was excluded from the sample analysis with the patient remaining evaluable.)

Criteria for Effectiveness and Analyses

The effectiveness of the Salest™ test was based on its clinical performance as a risk assessment marker for spontaneous PTL and PTD. A number of mutually exclusive patient categories were defined for purposes of statistical analyses. The test variable considered in all analyses was the highest intra-patient concentration of E3 collected between 23 to < 36 weeks of gestation. In addition to the tests delineated in the statistical plan, traditional statistical measures of test performance were also to be presented such as sensitivity, specificity, positive predictive values, and negative predictive values.

The primary clinical endpoint compared the difference in the incidence of PTL resulting in a PTD between patients who had a high E3 level (≥ 2.1 ng/mL) to those who had a low level (< 2.1 ng/mL) prior to 36 weeks gestation. The data were analyzed using Fisher's Exact test (2-tailed), with alpha = 0.05. The group of primary interest was women not treated with Betamethasone since this drug is known to suppress E3 levels and confounds the test results. The primary groups analyzed were PTL with PTD and term delivery (TD) with no PTL. Also the ability of E3 to predict PTL to traditional Creasy scoring was compared. Disagreements between the two methods were analyzed using McNernar's test (2-tailed), with alpha = 0.05. Additionally, the effectiveness of single readings versus consecutive (re-screen) E3 reading ≥ 2.1 ng/mL were also evaluated.

Despite the use of separate protocols for assessing Salest™ test performance in high- and low-risk patients, no statistically significant difference in performance between the two groups was expected based on results of the feasibility study. If afterward it was determined no statistically significant difference existed between the two groups, the two groups were combined and statistical analyses performed. Use of the Mantel-Haenszel test was planned to substantiate the combining of the two groups.

Determination of Sample Size

A power analysis was performed to determine the sample size needed to detect a difference in the incidence of PTL resulting in a PTD between women with a high E3 level (≥ 2.1 ng/mL) to those with a lower E3 level (< 2.1 ng/mL). Estimates were made for the low and high Creasy score patient samples separately. Both were based on the results from a multi-center feasibility study. Estimates for the low Creasy score sample were based on the 67 patients (single-fetus) classified as such. Estimates for the high Creasy score sample were based on the results of the 160 patients (single-fetus) classified as such. For purposes of estimating sample size, tables 5 and 6 were prepared using the results from the feasibility study.

Table 5
Feasibility Study Results - Low-Risk Patients

Number of Patients	PTL with Preterm Delivery	No PTL with and without PTD and PTL with Term Delivery	Totals
E3 < 2.1 ng/mL	1	35	36
E3 ≥ 2.1 ng/mL	6	25	31
Totals	7	60	67

Table 6
Feasibility Study Results - High-Risk Patients

Number of Patients	PTL with Preterm Delivery	No PTL with and without PTD and PTL with Term Delivery	Totals
E3 < 2.1 ng/mL	5	99	104
E3 ≥ 2.1 ng/mL	10	46	56
Totals	15	145	160

The following specifications and assumptions were used to estimate sample size:

- 2-tailed significance level of 0.05
- power of 80%
- difference to be detected 5% to 20% for an effect size of approximately 0.5

For the high Creasy scoring patient group a total sample size of approximately 150 evaluable patients was estimated to be needed to detect a significant difference at the alpha = 0.05 level. To account for dropouts and noncompliance, 200 patients were to be enrolled into the study. This number was increased to 300 to be reasonably sure of assessing an adequate number of preterm deliveries.

For the low Creasy scoring group a total sample size of approximately 110 evaluable patients was estimated to be needed to detect a significant difference at the alpha = 0.05 level. To account for dropouts and non-compliance, 200 patients were to be enrolled into the study.

Of the total 500 patients, 400 were expected to be suitable for analysis with an expected yield of approximately 40 preterm deliveries.

Sample Processing

Processing of patient samples was initiated in October 1996.

Study Period

This study started in May 1995. Patient enrollment for the high-risk and low-risk study groups was stopped July1996. Samples were batched, and blinded testing began November 1996. The last patient delivered the following month. The database closed in February 1997.

A list of the investigators who participated in this study group is in Table 7. For future reference clinical site's will be identified by the corresponding site number.

Table 7 Investigators

01 Dr. James McGregor University of Colorado	02 Dr. Jay Iams The Ohio State University
[redacted]	
05 Dr. Raul Artal State University of New York	06 Dr. Patricia Robertson University of California San Francisco
[redacted]	

Patient Accounting

Primary Study Group

The study groups presented in support of the safety and effectiveness of the Salest™ test were comprised of both high and low-risk women who (1) underwent serial E3 monitoring during their pregnancy (starting at 21 weeks gestation and less than week 25, until delivery) and (2) were not treated with Betamethasone or tocolytics during the course of their participation. Since the great majority of patients given Betamethasone were also given tocolytics (only 6 patients were given tocolytics without Betamethasone) the primary data set used for analysis excluded patients treated with either Betamethasone or tocolytics.

In order to provide a comprehensive analysis of all patient data within the study, analyses performed on patients not treated with Betamethasone or tocolytics were repeated on all patients irrespective of treatment status. Data analyses are presented separately for the low- and high-risk patient subgroups as well as the total study group.

A. Rationale For Excluding Women Treated With Betamethasone/Tocolytics

Prior to initiation of the Pivotal Study it was recognized from the scientific literature that administration of Betamethasone likely would have a suppressing effect on E3 levels.

Henderschott and Goodwin[6] studied the effect of Betamethasone on E3 levels on 10 patients given weekly treatment with Betamethasone.

All specimens were analyzed by the Salest™ test. The results of this study showed an average drop of 23.1 % from pre- to post- Betamethasone levels but rebounding to the starting level prior to the next dose. When weekly pretreatment values were looked at over time, the geometric mean of the individual patients' slopes did not differ significantly from no change. The same was true of post treatment values.

A second study used data for Betamethasone treated women in the Biex Pivotal Study and did not for the most part involve repeated treatments. Data used for the analysis were from the first treatment and used the highest E3 values in the two weeks before treatment and the highest E3 values in the two weeks (approximately) after treatment. The ratios of post to pre-treatment E3 levels were calculated. The geometric mean ratio was 0.71 with a C.V. of 0.45. A paired t-test gave a $t = 7.16$ ($df = 60$) which showed that 0.71 differed from 1 (p very close to 0).

Results of the above two studies were similar and demonstrated that Betamethasone lowered saliva E3 levels by close to 30% on the average. Additionally, with a such large C.V.s as those in the two studies, it would be impractical to try to arrive at an "adjustment" for E3 levels in patients given Betamethasone. It was clear that Betamethasone had a substantial suppressing effect on salivary E3 levels. Consequently, the Salest™ test results can not be relied upon when testing patients in whom Betamethasone has been administered. Moreover, a decision to treat with this drug is usually based on an *a priori* diagnosis of PTL, therefore, continued use of the Salest™ test would be inappropriate. As noted previously, the analyses presented in support of the safety and effectiveness of the Salest™ test only included patient data exclusive of treatments with Betamethasone or tocolytics. The product labeling includes appropriate limitation statements regarding the suppressing effects of these drugs and their impact on test results.

B. Justification For Combining Low- And High-Risk Women

A Mantel-Haenszel Chi-Square Test (M-H Test) was applied to assess the result of combining the evidence from the low and high-risk patient subgroups. Further, a Woolf's test for heterogeneity was performed to see whether the odds ratios were significantly different for the two subgroups. The groups selected for this analysis were all patients with PTL and PTD within 72 hours and all patients with no PTL and PTD (excluding patients administered Betamethasone and/or tocolytics).

Combining the evidence using the Mantel-Haenszel procedure showed a highly significant difference in the proportions of positive and negative cases, $p = 0.002$. Further, Woolf's test for heterogeneity revealed that the odds ratios were not significantly different ($p = 0.8930$). It appeared justifiable to combine the evidence of the low and high-risk patient subgroups and evaluate the group as one.

PRIMARY STUDY FINDINGS

The results of the Pivotal Study demonstrated the following salient points:

- *Primary Endpoint Data and Analysis:* The incidence of spontaneous PTL and PTD was statistically significantly increased when a single salivary E3 value ≥ 2.1 ng/mL was obtained prior to 36 weeks in the study group (high and low-risk women together).
- In low-risk women elevated salivary E3 levels identified 50 % of spontaneous PTL and PTD cases that would have otherwise been missed by application of traditional risk factors (Creasy).
- In high-risk women an elevated salivary E3 level was a more accurate predictor of PTL and PTD than traditional risk factors (Creasy).
- A second elevated salivary E3 test (re-screen) following an initial elevated salivary E3 result (i.e., 2 consecutive positive tests and women who delivered prior to the second test) enhanced the risk assessment properties of salivary E3 as a stand-alone risk assessment marker.
- Two consecutive positive Salest™ tests (re-screen test) resulted in 70% of the patients delivering within 3 weeks (independent of gestational age and treatment). The re-screen in PTL and PTD patients resulted in 87% of the patients delivering in less than 2 weeks and 100% in less than 3 weeks.
- The negative predictive value of the single and re-screen test in predicting no spontaneous PTL and PTD was 98% (untreated women). Use of a re-screen with a low E3 value resulted in an incidence of spontaneous PTL and preterm birth of 2%. Incidence of spontaneous PTL and birth with a high E3 value was 19% (untreated women).
- The results of the present study are consistent with the findings of the earlier feasibility study. These results can be examined together using a meta-analysis.

The clinical study results and analyses provided in support of the above findings and conclusions are presented in the same order as above. For purposes of completeness, data presented for the total study group is followed by the data for the low and high-risk patient subgroups separately.

1. Primary Endpoint Data and Analysis

In the Pivotal Study there were 601 evaluable patients. Tables 8-10 provide specific performance characteristics of the Salest™ test and corresponding 2 x 2 contingency tables for the total study group, as well as the low and high-risk patient subgroups. For purposes of the primary analysis, the incidence of PTL with PTD for patients who had an E3 ≥ 2.1 ng/mL is compared to the incidence in those patients who had an E3 < 2.1 ng/mL.

Table 8
Primary Endpoint Contingency Table
Total Study Group (N=601)

	PTD	TD	Totals
< 2.1	10	449	459
≥ 2.1	13	129	142
Totals	23	578	601

Performance Characteristics

Fisher's Exact Test: p = <0.001*
Sensitivity = 56.5% (34.5-76.8)**
Specificity = 77.7% (74.1-81.0)
NPV = 97.8% (96.0-99.0)
PPV = 9.2% (5.0-15.2)
Relative Risk = 4.20 (1.88-9.38)
Incidence + E3 (Incidence of Positivity) = 23.6%
Incidence of PTL with PTD = 3.8%
M-H test for combined evidence: p=0.002
*Significant at the 0.05 level **95% binomial confidence limits are presented in parentheses

Table 9
Primary Endpoint Contingency Table
Low-Risk Patients (N=449)

	PTD	TD	Totals
< 2.1	6	353	359
≥ 2.1	6	84	90
Totals	12	437	449

Performance Characteristics

Fisher's Exact Test: p= 0.018*
Sensitivity = 50.0% (21.1-78.9)**
Specificity = 80.7% (76.7-84.4)
NPV = 98.3% (96.4-99.4)
PPV = 6.7% (2.5-14.0)
Relative Risk = 3.99 (1.32-12.08)
Incidence + E3 (Incidence of Positivity) = 20.0%
Incidence of PTL with PTD = 2.7%
*Significant at the 0.05 level **95% binomial confidence limits are presented in parentheses

Table 10
Primary Endpoint Contingency Table
High-Risk Patients (N=152)

	PTD	TD	Totals
< 2.1	4	96	100
≥ 2.1	7	45	52
Totals	11	141	152

Performance Characteristics

Fisher's Exact Test: p= 0.047
Sensitivity = 63.6% (30.8-89.1)
Specificity = 68.1% (59.7-75.7)
NPV = 96.0% (90.1-98.9)
PPV = 13.5% (5.6-25.8)
Relative Risk = 3.4 (1.03-10.97)
Incidence + E3 (Incidence of Positivity) = 34.2 %
Incidence PTL with PTD = 7.2%

Discussion and Conclusions

The incidence of spontaneous PTL and PTD was statistically significantly increased when the level of salivary E3 was ≥ 2.1 ng/mL prior to 36 weeks in the total study group (high- and low-risk women together).

The statistics above showed that the Salest™ test detected over half of the PTL with PTDs in the study group, high- and low-risk subgroups together. Moreover, the relative risk of a woman with a high E3 was 4.2, thus a women with a high E3 was 4.2 times as likely to have a PTL with PTD than a woman with a low E3.

In light of the above salivary E3 when used as a single test was an independent risk assessment marker for PTL with PTD. This finding was based on the results of the patient having at least one positive test result (E3 ≥ 2.1 ng/mL).

2. The results of the Primary Endpoint Analysis for Low-Risk Patient Subgroup/Traditional Risk Factors (Creasy) are presented above.

Discussion and Conclusions

The Salest™ test detected 50% of the PTLs with PTDs in the low-risk group with a corresponding relative risk of 3.99. That is, of the 12 PTL with PTD that occurred in the low-risk group defined by the Creasy system, the Salest™ test identified 6 of these (Sensitivity = 50%) for a 50% increase in detection over traditional risk factors. The performance of the Salest™ test in this group was noteworthy in that, as previously stated, 50% of women who deliver preterm are determined to be low-risk based on traditional risk factors. A number of women with PTL with PTD could be detected by the Salest™ test who otherwise were considered low-risk by traditional risk assessment criteria.

3. Primary Endpoint Analysis for High-Risk Patient Subgroup/Traditional Risk Factors (Creasy)

In the high-risk subgroup, a comparison was made between the number of patients classified by the Creasy system who were incorrectly classified by the Salest™ test and the number of patients classified by the Salest™ test who were incorrectly classified by the Creasy system. A McNemar's test (2-tailed), with alpha = 0.05 was used for this evaluation. Table 11 presents the results of this analysis.

Table 11
Resulting Contingency Table
For McNemar Test
Salest™ /Creasy Comparison

	Creasy Risk Assessment		
Salest™	Correct	Incorrect	Total
Correct	7	96	103
Incorrect	4	45	49
Total	13	141	152

McNemar Test Results
Chi-square: p<.001
Odds Ratio = 24.00 (95% CI: 8.83-65.25)

Result of the analysis using the Creasy high-risk (n = 152) subgroup showed a difference in the proportions of patients where the Creasy was incorrect and E3 was correct, versus where the Creasy was correct and E3 was incorrect (p<0.001).

Discussion and Conclusions

In a group defined as high-risk by traditional risk factors, the Salest™ test was more capable of accurately predicting the risk of PTL with PTD than traditional risk factors inherent in Creasy Scoring. Results of the analysis showed a difference in the proportions of patients where Creasy was correct and E3 was incorrect (2.6%) versus Creasy was incorrect and E3 was correct (63.2%) (p <0.001). Moreover, there was a 24-fold increase in favor of the Salest™ test in correctly predicting PTL with PTD over the Creasy score.

4. Re-screen Test Results

The use of a re-screen test in women with a single positive E3 test before 36 weeks of gestation was evaluated. The re-screen test was also performed before 36 weeks gestation. The purpose of the re-screen was to determine whether a repeat test enhanced a patient's risk profile relative to a single elevated E3 value. Several factors were analyzed, including the incidence of PTL with PTD and the time to delivery from the second consecutive positive E3 result. In the high-risk subgroup, results of the re-screen were also compared to those of traditional risk factors. The re-screen analysis examined the effects of two consecutive positive Salest™ test results during the course of serial

monitoring of E3 prior to 36 weeks of gestation. Shown below are the performance characteristics illustrating the results of these analyses in the total group and in the high- and low-risk subgroups.

(A) Enhanced Prediction of PTL and PTD

The effectiveness of the re-screen was evaluated using the primary endpoint of the study. That is, the incidence of PTL/PTD < 37weeks given two consecutive high E3 (≥ 2.1 ng/mL) tests was compared to the incidence without two consecutive high values. (The analysis included patients who had a single elevated E3 and delivered within a week, prior to the opportunity to have a second E3 test.)

Total Study Group

Performance Characteristics (N=601)

Fisher's Exact Test: P< 0.001*
Sensitivity = 43.5% (23.2-65.5)**
Specificity = 92.4% (89.9-94.4)
NPV = 97.6% (96.0-98.7)
PPV = 18.5% (9.3-31.4)
Relative Risk = 7.79 (3.59 - 16.92)
Incidence + E3 (Incidence of Positivity) = 9.0%
Incidence PTL with PTD = 3.8%
M-H Test for combined evidence: P <0.000001*
*Significant at the 0.05 level
**95% binomial confidence limits are presented in parentheses

Low-Risk Patient Subgroup (N=449)

Fisher's Exact Test: p< 0.001*
Sensitivity = 41.7% (15.2-72.3)**
Specificity = 93.1% (90.3-95.3)
NPV = 98.3% (96.6-99.3)
PPV = 14.3% (4.8-30.3)
Relative Risk = 8.45 (2.83-25.24)
Incidence + E3 (Incidence of Positivity) = 7.8%
Incidence PTL with PTD = 2.7%
*Significant at the 0.05 level
**95% binomial confidence limits are presented in parentheses

High-Risk Patient Subgroup (N=152)

Fisher's Exact Test: P= 0.005*
Sensitivity = 45.5% (16.8-76.6)
Specificity = 90.1% (83.9-94.5)
NPV = 95.5% (90.4-98.3)

PPV = 26.3% (9.2-51.2)
Relative Risk = 5.83 (1.97-17.27)
Incidence + E3 (Incidence of Positivity) = 12.5%
Incidence PTL with PTD = 7.2%
*Significant at the 0.05 level

Discussion and Conclusions

The study demonstrated that use of a re-screen test following an initial positive result further enhanced risk assessment results with a $p < 0.001$ for the total study group, $p < 0.001$ for the low-risk subgroup and $p < 0.005$ for the high-risk subgroup. Moreover, overall elevated E3 carried a relative risk of 7.79 for the total group, 8.45 for the low risk subgroup and 5.83 for the high-risk subgroup compared to traditional risk factors such as twins or prior preterm birth which have a reported relative risk of 2.0 to 2.5. Further, less than 10% of the group was identified as at risk, minimizing false positives.

(B) Enhanced Prediction of PTL and PTD Over Traditional Risk Factors

A comparison was made between the number of patients in the total group classified by the Creasy system who were incorrectly classified by the Salest™ re-screen test and the number of patients classified by the Salest™ re-screen test who were incorrectly classified by the Creasy system. A McNemar's test (2-tailed), with alpha = 0.05 was used for this evaluation.

Table 12
Contingency Table
Salest™ Re-screen/ Creasy Comparison (Total Group)

	Creasy Risk Assessment		
Salest™ Rescreen	Correct	Incorrect	Total
Correct	412	132	544
Incorrect	36	21	57
Total	448	153	601

McNemar Test Results
Chi-square: $p<0.001$
Odds Ratio = 3.67 (95% CI: 2.54 - 5.3)

Result of the analysis using the Creasy high-risk (n = 152) and low-risk (n = 449) subgroup showed a difference in the proportions of patients where Creasy was incorrect and E3 was correct versus Creasy was correct and E3 was incorrect ($p < 0.001$).

Table 13 presents a similar analysis for the high-risk group only.

Table 13
Contingency Table
Salest™ Re-Screen/ Creasy Comparison (High Risk)

	Creasy Risk Assessment		
Salest™ Rescreen	Correct	Incorrect	Total
Correct	5	127	132
Incorrect	6	14	20
Total	11	141	152

McNemar Test Results
Chi-square: $p<0.001$
Odds Ratio: 21.17 (95% CI: 9.33 - 48.00)

Result of the analysis using the Creasy high-risk subgroup (n=152) showed a difference in the proportion of patients where Creasy was incorrect and E3 was correct versus Creasy was correct and E3 was incorrect ($p<0.001$).

Discussion and Conclusions

In the total group, the Salest™ test was more capable of accurately predicting the risk of PTL with PTD than traditional risk factors inherent in Creasy Scoring. Results of the analysis showed a difference in the proportions of patients where Creasy was correct and E3 was incorrect (6.0%) versus Creasy was incorrect and E3 was correct (22.0%) ($p<0.001$). Moreover, there was a 3- fold increase (based on Odds Ratio) in favor of the Salest™ test in correctly predicting PTL with PTD over the Creasy score.
In the high-risk group, the Salest™ test was a better predictor of risk of PTL and PTD than traditional risk factors. Results of the analysis showed a difference in the proportion of patients where Creasy was correct and E3 incorrect (3.9%) versus Creasy incorrect and the Salest™ test correct (83.6%) ($p < 0.001$). Moreover, there was a 21.17 fold increase (based on Odds Ratio) in favor of the Salest™ test in correctly predicting PTL with PTD over the Creasy score in this group.

(C) Time to Delivery

The accuracy of the Salest™ test in predicting delivery within 1 to 5 weeks of either a single positive test or two consecutive positive tests (re-screen test) was examined. Two evaluations were performed, the first on all 714 evaluable patients (irrespective of term or PTD) and the second on only those patients who had PTL with PTD.

Discussion and Conclusions

The results demonstrated that a single positive salivary E3 value was capable of predicting the probability of PTL and PTD within a 5-week time frame, regardless of

gestational age. Moreover, use of the re-screen test enhanced the accuracy with which the salivary E3 could have predicted the probability of delivering within this same 5 week time frame. In the subpopulation of women who had preterm labor and delivery, women who had a positive rescreen had a 63% chance of delivering within 1 week, an 88% chance of delivering within 2 weeks, and a 100% chance of delivering within 3 weeks. The data on all evaluable patients demonstrated that a single positive Salest™ test indicates a 78% likelihood of delivering within 5 weeks compared to a 92% likelihood of delivering within 5 weeks after a positive re-screen.

(D) Length of Time No delivery

The accuracy of the SalEst™ test in predicting **no delivery** was also examined for the 4-week period following each negative SalEst™ test (E3 <2.1 ng/mL). The analysis was based on a subgroup of all evaluable patients who met the criteria described above.

Discussion and Conclusions

Although it was evident that an elevated E3 was indicative of a women's risk for PTL and PTD, the clinical utility of the Salest™ test was not necessarily limited to "positive" test results. A useful risk assessment marker for PTL and PTD should also provide information about marker levels that are not elevated. A negative Salest™ test result predicted the likelihood of not delivering within the ensuing 2 weeks. At early gestational ages (i.e.,< 32 weeks gestation), the probability that PTL and PTD will occur during the 4 weeks following a negative E3 test was no greater than 3%. For example, when a sample collected at 32 weeks was negative, the woman had a 98% chance of not delivering within the next week, and a 97% chance of not delivering before week 34.

ADDITIONAL EFFECTIVENESS ANALYSES

Meta-Analysis

For purposes of an overall evaluation of the Salest™ test clinical data, a meta-analysis was performed. A Mantel-Haenszel Chi-Square Test (M-H Test) was used to assess the appropriateness and result of combining the evidence from the feasibility and pivotal studies.

The analysis was performed on the results for the total group as well as the two risk subgroups. As shown in Table 14, the data from the two studies were consistent with one another.

Table 14

Primary Endpoint Analysis
Combined Contingency Table
Feasibility and Pivotal Studies
Single Test

Study	Positive E3 Cases		Negative E3 Cases	
	PTL	No PTL	PTL	No PTL
Feasibility High-risk	4	41	5	102
Feasibility Low-risk	6	18	1	35
Pivotal High-risk	7	45	4	96
Pivotal Low-risk	6	84	6	353

Geometric Means of Estriol by 2-Week Intervals

Table 15 presents the geometric means of the E3 levels by gestational 2-week intervals (excluding all patients treated with either Betamethasone and/or tocolytics) for patients with PTL and PTD and patients with no PTL who delivered at term. The highest intra-patient E3 value recorded within the interval was used to calculate the average value. Using a t-test, the means at each gestational week were compared. The mean E3 level of patients with PTL and PTD was higher at all intervals and statistically significantly higher ($p < 0.05$) at weeks 30 to <32, 36 to <37, and 30 to <37.

Table 15

E3 Levels (Geometric Means)
By Gestational Age, Weeks

Age, Week	**PTL with PTD**	**No PTL with Term Delivery**	**P-value**
20 - < 22	0.72 (n = 4; 0.45 - 1.15)	0.61 (n = 121; 0.42 - 0.89)	0.403
22 - < 24	0.79 (n = 16; 0.51 - 1.20)	0.74 (N = 357; 0.51 - 1.06)	0.499
24 - < 26	0.96 (n = 23; 0.74 - 1.24)	0.85 (n = 564; 0.61 - 1.18)	0.081
26 - < 28	1.06 (n = 23; 0.81 - 1.37)	0.93 (n = 572; 0.67 - 1.30)	0.076
28 - < 30	1.09 (n = 22; 0.66 - 1.78)	1.02 (n = 568; 0.73 - 1.42)	0.561
30 - < 32	1.29 (n = 23; 1.07 - 1.55)	1.14 (n = 570; 0.81 - 1.61)	0.005
32 - < 34	1.43 (n = 23; 1.05 - 1.94)	1.30 (n = 575; 0.94 - 1.80)	0.173
34 - < 36	2.01 (n = 23; 0.94 - 4.29)	1.61 (n = 572; 1.07 - 2.42)	0.183
36 - < 37	2.70 (n = 9; 1.76 - 4.13)	1.77 (n = 521; 1.15 - 2.73)	0.004
30 - < 37	2.30 (n = 23; 0.88 - 3.73)	1.72 (n = 578; 0.55 - 2.90)	0.009

Conclusions Drawn from the Studies

Based on the above non-clinical and clinical studies the Salest™ test is a safe and effective risk assessment marker for detection of spontaneous preterm labor and preterm delivery (PTL and PTD) in women with singleton pregnancies. Additionally, it is a strong negative predictor for PTL and PTD in women with low Salest™ levels.

IX. Panel Recommendation

On December 10, 1997, the FDA Clinical Chemistry and Clinical Toxicology Devices Panel recommended that Salest™ be approved with conditions. The following are recommendations made by the Panel:

1. Submission of demographic data for patients who tested positive with the test.
2. Exploring the predictive values at different cut-off points.
3. Addressing frequency of testing, variance around the 2.1 ng/mL cut-off, and serial testing and its effect on the predictive values.
4. Revising the labeling accordingly including condensing the physician's monograph.

X. CDRH Action on the Application

CDRH concurred with the recommendations of the Panel and issued an approvable letter to the applicant on January 21, 1998 requesting the information. In that letter, CDRH also requested that certain patients be removed from the data set as well as some samples from individual patients, and that the performance data be recalculated. The sponsor responded to the January 21, 1998 letter in an amendment, which was received on February 17, 1998.

The applicant's manufacturing and control facilities were inspected March 18 through March 24, 1998 and the facilities were found to be in compliance with the Good Manufacturing Practice Regulations (GMPs). The shelf-life of the Salest™ has been established at one year stored at 2-10°C. CDRH issued an approval order on April 29, 1998.

XI. Approval Specifications

Directions for use: See attached labeling

Conditions of Approval: CDRH approval of this PMA is subject to full compliance with the conditions described in the approval order dated April 29, 1998.

REFERENCES

1. Creasy RK, Herron, MA. Prevention of Preterm Birth. Seminars in Perinatology July 1981; 5(3):295-302.

2. Danforth's Obstetrics and Gynecology, Sixth Edition, Scott, JR.; J.B. Lippincott Company. Philadelphia, 1990:70-71.

3 ACOG Technical Bulletin, Preterm labor. Number 133, October 1989.

4 B.M. Mercer et. al., The preterm prediction study: A clinical risk assessment system, Am. J. Obstet Gynecol. 1996;174(6):1885-95.

5 Centers for Disease Control. Increasing incidence of low birthweight - United States 1981-1991. MMWR Morb Mortal Wkly Rep 1994; 43:335-9.

6 C.M. Henderschot, V. Dullien, T.M. Goodwin. Dept. Obstetrics and Gynecology, LAC+USC Medical Center, Women's and Children's Hospital, Los Angeles, CA