

Statistical Review of Lifecore's Clinical Study

Review of Study Protocol

I have attached the sponsor's study protocol (Attachment A, dated 1/26/95) for your review. Basically, it states the following:

- 1) the pilot study (see Attachment A, page 34 of 38, under section H. Power) found a difference of adhesion score of 4.0 between treated and control groups (mean adhesion scores of 1.7 and 5.7, respectively, with corresponding standard deviations of 1.4 and 2.7, respectively);
- 2) the sponsor agreed to do an intent-to-treat analysis with the lost to follow-up being treated as failures and getting the worst score (a 16 according to their scoring system);
- 3) the sponsor expected an unequal lost to follow-up for the 2 groups: 20% lost to follow-up for the treated group and 10% for the control;
- 4) they used an intent-to-treat (ITT) analysis extrapolated from the pilot study with the above lost to follow-up rates (20% for Intergel and 10% for control treated patients) which yielded a mean adhesion score of 4.6 (st. dev. of 5.9) for the Intergel patients and 6.7 (standard deviation of 4.1) for the control group;
- 5) which results in a total of 180 patients necessary to detect, with 80% power at the $\alpha=0.05$ significance level, a difference in adhesion score of 2.0 given an expected standard deviation of 5.0 for both groups; and
- 6) due to the skewness of the data (patients scores can range from 0 to 16 with an average around observed average adhesion score of 2 to 3 and with an observed standard deviations of 3.5 to 4.5), the sponsor said they would perform a nonparametric analysis.

The sponsor requested they be able to perform the study in both the US and Europe. They would perform an interim analysis to determine if 120 US and 80 European subjects were combinable to obtain a total of 200 subjects. If they were not combinable by the procedures laid out in the attached protocol (page 31 to 33 of 38, Attachment A), they would continue the US study and use a total of 200 US subjects (100 per arm) for their analysis and not use the European data.

Review of Study Results

We looked at the interim data from the US and European studies and determined that the subjects were not combinable. The baseline adhesion scores and incidence of adhesion scores were very different (see Tables 1 and 2 below). The difference in change from baseline for the 2 groups was also large. The US patients had a doubling to tripling of their baseline score at 2nd look while the European patients score changed much less from baseline to 2nd look. Also the type of patients differed greatly. The U.S. had mostly non-adhesiolysis patients while the Europeans were mostly adhesiolysis patients. The sponsor disagreed with our conclusion that the two groups were different but continued to enroll

patients until they had enrolled 200 US patients. They had also completed their complement of 81 European patients. Though the sponsor sized the trial for 180 to 200 patients, they based their analyses on all 281 U.S. and European patients.

The data in Table 1 below presents the intent-to-treat (ITT) summary statistics for the mAFS score and number of adhesions at baseline and 2nd look adjusted (for baseline adhesions not lysed) for both the patients treated with Intergel (IG) and the control, Lactated Ringer' Solution. These summary statistics were provided by the sponsor and extracted from the back of their December 1999 Panel Pack (Panel Pack II, Data 1, page 1) for the General and Plastic Surgery Advisory Panel.

Table 1. Intent-to-Treat Patient Population

	United States			Europe		
	Intergel N=102	Control N=98	Difference (IG-C)	Intergel N=41	Control N=40	Difference (IG-C)
MAFS score						
Baseline	0.78	0.68	+0.10	1.57	1.95	-0.38
2 nd Look Adj	2.63	2.76	-0.13	2.01	2.12	-0.11
Number of Adhesions						
Baseline	2.49	2.27	+0.22	6.00	6.40	-0.40
2 nd Look Adj	7.92	7.73	+0.19	6.63	7.33	-0.70

Note that there are no statistically significant results between the 2 groups for both mAFS and number of adhesions for the US patients or the European patients. In fact we see no apparent difference between the treatment and control groups. (Remember in the pilot study we saw a difference of 4.0 and, after accounting for lost to follow-up as specified in the protocol, the study was designed to detect a difference of 2.1) Thus as designed, the trial does not show the device to be effective. In fact, all the differences we see in this table are very small, except for the difference in the baseline scores between the European and US patients and the difference in change from baseline to 2nd look adjusted between the US and European patients. European baseline scores are 2 to 3 times larger than the corresponding US scores. The US patients also have a 3- to 4-fold increase over baseline in both adhesion score and number of adhesions while the European patients have a very small increase (of only 10% to 30%) over baseline. Thus the patients from the 2 continents are very different at baseline and appear to respond very differently to treatment with respect to adhesion and mAFS score indicating the 2 groups are not combinable. Therefore, it was determined that the 200 US patients comprised the appropriate patient group to analyze for device effectiveness.

The sponsor presented the summary results based on evaluable patients as opposed to the intent-to-treat patients. Their analysis violated many of the

premises upon which the study was designed (as stated above in the protocol section):

- 1) The sponsor used only evaluable patients (which excludes those patients lost to follow-up) instead of including lost to follow-up patients as described in the study protocol.
- 2) The sponsor used all 265 evaluable patients from the US and Europe (which is greater than the 200 US patients allowed or the 180 patients for which the study was sized).
- 3) Since the patients lost to follow-up were removed from the analysis, the observed standard deviations for the mAFS scores (of 1.5 and 2.6 for the Intergel and control groups, respectively) are much smaller than the expected intent-to-treat standard deviation (of 5.0) for which the study was designed.

All these conditions lead to a vastly overpowered study that could result in finding statistically significant differences between groups that may not be clinically meaningful. So, if it were appropriate to combine patients across continents and use evaluable patients only, the test would provide 80% power to detect a difference of only 0.75 difference in the mAFS. This is much smaller than the agreed upon 2.1 clinical difference in mAFS score that the ITT study was designed to detect.

In summary, using the ITT design and analysis presented by the sponsor in their study protocol, there is no statistical difference between the Lactated Ringer's and control groups with respect to mAFS score or adhesion score at second look.

Shift Tables for AFS-scores

Next, consider the new post-hoc analysis that the sponsor has submitted in their post-panel meeting PMA amendment (P990015/A10). This report presents shift tables for the American Fertility Scoring System (AFS) which scores adnexal adhesions only and that these scores were obtained retrospectively by a method which approximates the AFS scoring system. Also, note that this data was presented by the sponsor to the experts at the January 2000 Advisory Panel meeting (Attachment C) at which the Panel concluded that the product did not provide a clinically meaningful benefit. In this Table (in Attachment C) and in their analysis the sponsor presents data from all evaluable patients in both continents but ignore those patients lost to follow-up. Again, it should be emphasized that the study was not designed to analyze AFS data and that the sponsor is performing post-hoc analyses on data that was already presented to the Panel.

Table 3 (below) is the ITT presentation of the shift table for sponsor's (retrospectively calculated) AFS data for US patients. (I had to combine the first two categories none and minimal and mild since intent-to-treat shift tables, stratified by continent, were not provided by the sponsor.) The denominator is the number of subjects in each subgroup (none/minimal/mild and moderate/severe) having the baseline adhesion status specified by that subgroup. The numerator is the number of subjects whose adhesion status at 2nd look is moderate/severe. Exploratory analysis of the retrospective data in Table 3 found no statistical differences between the treatment groups whether analyzed by subgroup or as a whole.

TABLE 3.

Number of Patients with Moderate or Severe AFS Scores (>10) at Secondlook: Intent-to-Treat Patients

<u>Adhesion Status at Baseline</u>	<u>Intergel</u>	<u>Control</u>	<u>p-value*</u>
None, Minimal or Mild	12/97	11/91	0.99
Moderate or Severe	0/5	3/7	0.23

* p-value based on Fisher's Exact test for comparison of 2 proportions

In the sponsor's analysis in Amendment 11, they present an imputation scheme to account for the patients lost to follow-up instead of their original intent-to-treat analyses for which the study was designed. Their post-hoc method relies on deleting data from patients who did not have 2nd looks and did not have any complaints. Note that this method is not appropriate, nor can it be statistically justified because there is no way of knowing how the patients without complaints or who did not return really fared. Furthermore, their method discards data from 8 of 12 of those Intergel patients lost to follow-up group, while only deleting 1 of the 4 control patients who were lost to follow-up; this approach biases the results in favor of the sponsor.

Surgical Site and Reformed Adhesions

In their PMA Amendment 11, the second part of their proposed *Indication for Use* states "INTERGEL Solution was also shown to reduce adhesion reformation to sites in addition to adnexa; and adhesion formation at surgical sites, including the anterior abdominal incision." In the discussion and analysis of adhesion reformation and surgical site adhesions presented on page 46-7 of Section III of Amendment 11, they only present evaluable data for all US/European patients combined (ignoring the appropriate intent-to-treat analyses stratified by continent). Furthermore, after having failed to show effectiveness of the primary study endpoint, the sponsor has chosen two of several secondary endpoints (surgical site and reformed adhesions) for which they claim Intergel is superior. Therefore, Amendment 11 proposes the situation where, not only is the appropriate intent-to-treat analysis stratified by continent discarded, but a couple of several secondary endpoints defined in the original study are evaluated without any adjustment of the significance level of the statistical tests. A proper multiplicity adjustment is required to lower the significance level of the tests to adjust for the multiple subgroups. In addition, clinically meaningful differences for these endpoints were not defined *a priori*, and thus, the study was neither designed nor powered to assess them in a statistically valid fashion. The intent-to-treat results for these endpoints (reformed and surgical site adhesions), as well as denovo adhesions, are presented in Table 4 below. These summary statistics were provided by the sponsor and extracted from the back of their December 1999 Panel Pack (Panel Pack II, Data 1, page 1) for the General and Plastic Surgery Advisory Panel. No statistical differences were found for any of these endpoints, even without a multiplicity adjustment.

Table 4. Secondary Endpoints: Intent-to-Treat, U.S. Patients

Mean Incidence	United States		
	Intergel N=102	Control N=98	Difference (IG-C)
# adhesions lysed*	2.09	1.89	+0.20
Reformed Adhesions	3.18	3.51	-0.38
DeNovo Adhesions	6.71	6.34	+0.37
Surgical Site Adhesions	2.29	2.62	-0.33

*Average incidence at baseline minus average incidence after 1st surgery

Conclusion

Using the statistical analysis plan from the study protocol (intent-to-treat analysis on the US patients), the sponsor was unable to demonstrate that patients treated with Intergel had statistically lower mAFS score or statistically fewer adhesions than the control, Lactated Ringer's solution. In fact, in the United States, both products showed approximately equivalent increases from baseline in both adhesion incidence and mAFS score and these increases were substantial. Furthermore, the sponsor was unable to demonstrate that it is valid to combine the data across continents as both the baseline and change from baseline for both mAFS score and incidence of adhesions were very different for the Intergel and the control. The January 2000 General and Plastic Surgery Advisory Panel determined that there was not reasonable assurance that the product was safe and effective. In their subsequent PMA Amendment, the sponsor did not present any new data, but only selectively re-analyzed (in an unplanned, post-hoc fashion) data already presented at the earlier panel meeting.

ATTACHMENT A

SPONSOR'S STUDY PROTOCOL

X. STATISTICAL METHODS

XI. A. PATIENT POPULATIONS

1. The intent-to-treat efficacy and safety populations will consist of all patients who receive LUBRICOAT Gel or Lactated Ringer's Solution.
2. A subset of the intent-to-treat efficacy population will exclude patients who refuse the second-look laparoscopy for reasons unrelated to the device.
3. The evaluable efficacy population will consist of all patients who receive a second-look laparoscopic evaluation.

Patients who are randomized but do not receive treatment will be described but will not be otherwise analyzed. If any patients are incorrectly randomized, alternative analyses will be performed with those patients analyzed in the treatment group or the assigned group.

B. EFFICACY VARIABLES

The primary efficacy variable will be a total adhesion score using the Adhesion Scoring Method of the American Fertility Society (AFS) applied to 24 anatomical sites. Adhesions occurring at each of the 24 potential adhesion sites will be scored as mild (a filmy avascular adhesion) or severe (a dense organized cohesive vascular adhesion). The extent of adhesions will be graded as Localized (<1/3 of the site covered), Moderate (1/3-2/3 of the site covered) or Extensive (>2/3 of the site covered). The extent of adhesions will not be scored for the small bowel, omentum and left and right large bowel since their size precludes adequate visualization. These sites will be assigned a classification of Moderate in order to determine the total adhesion score.

For each adhesion site, the adhesion score will be derived from severity and extent scores as follows:

No Adhesion		
Severity: Mild	Extent: Localized	1
Severity: Mild	Extent: Moderate	2
Severity: Mild	Extent: Extensive	4
Severity: Severe	Extent: Localized	4
Severity: Severe	Extent: Moderate	8
Severity: Severe	Extent: Extensive	16

Scores from all potential adhesion sites will be averaged to yield a total adhesion score. Adhesions will be characterized as *de nova* if the site had no pre-existing adhesions and as reformed if the site had adhesions that were lysed during the original surgery. Sites with *de novo* adhesions will also be characterized as surgical versus non-surgical.

These analyses will be conducted for all sites as well as for pelvic and abdominal site groupings. Pelvic sites include the caudal anterior peritoneum, anterior and posterior uterus, cul-de-sac, right and left pelvic sidewall and all tube, ampulla and ovarian sites. Abdominal sites include the right and left cephalad anterior peritoneum, small bowel, omentum, right and left large bowel, rectosigmoid and the anterior peritoneum incision.

The proportion of sites with adhesions will be analyzed as a secondary efficacy variable. This will be a mean proportion based on the number of sites with adhesions divided by the number of possible adhesion sites. As above, adhesions will be characterized as *de nova* versus reformed, surgical versus nonsurgical, and pelvic versus abdominal.

In addition, adhesion sites will be categorized by the presence or absence of endometriosis, use of sutures and the method of adhesiolysis (sharp dissection, blunt dissection, cautery, laser). Each anatomical site will also be analyzed.

Additional secondary variables will include the extent and severity of all categories of adhesions. Severity will be scored on a three-point scale where 0 = None, 1 = Mild and 3 = Severe. Extent will be scored on a four-point scale where 0 = None, 1 = Localized, 2 = Moderate and 3 = Extensive.

C. SAFETY VARIARBLES

Safety variables will include the proportions of patients reporting adverse events categorized using COSTART terms. Laboratory values will be presented as mean change from baseline and as transition tables showing the proportions of patients above, below and within the normal range before and after treatment.

D. DEMOGRAPHIC, PRETREATMENT AND SURGICAL VARIABLES

Age, race, height, weight, blood pressure, previous and concomitant medications (categorized by AHFS codes), presence of endometriosis, surgical procedures (categorized by CPT codes), estimated blood loss, operative time, baseline adhesion scores and length of hospital stay will be analyzed. Use of these variables to determine combinability with a European study (Protocol PTL-0022) is described in Section G.

E. STATISTICAL ANALYSIS

Second-look adhesion scores will be analyzed using factorial analysis of covariance where one factor is treatment group (LUBRICOAT Gel versus Lactated Ringer's Solution), the other factor is center and baseline adhesion score is a covariate. This will allow analyses of the effect of treatment, the effect of center and the interaction of treatment with center. Homogeneity of slopes will be tested by examination of interactions between baseline adhesion score and treatment group.

If the two groups differ on any important demographic or surgical variables or if these pre-treatment variables appear to strongly predict second-look adhesion scores (as determined using multiple linear regression with treatment group forced into the model as a dummy variable), these variables may be added to the model as covariates. Homogeneity of slopes will be tested by examination of interactions between covariates and treatment group. Pretreatment variables may be transformed in order to yield homogeneous slopes.

The mean proportion of sites with adhesions at second look will be analyzed in the same fashion as the mean second-look adhesion scores.

Other continuous variables will be analyzed using factorial analysis of variance where one factor is treatment group and the other factor is center. Analyses to determine combinability will also use continent (US versus Europe) as a factor (see Section G).

Categorical variables will be analyzed using the Cochran-MantelHaenszel test with individual sites as strata. Determination of combinability of the US and European data will use categorical models as described in section G. Proportions with small expected event rates (e.g. adverse events) will be analyzed using Fisher's exact test. Laboratory value transition tables will be compared using 2x9 Fisher's exact tests.

Two-sided p values will be reported and p values less than 0.05 will be considered to be statistically significant.

F. INTENT-TO-TREAT ANALYSIS

As requested by FDA, an intent-to-treat analysis will be performed in which patients treated with LUBRICOAT Gel or Lactated Ringer's Solution who do not have a second-look laparoscopy will be considered to be treatment failures. This will be accomplished by assigning them second-look adhesion scores of 16 (the worst possible score). Because this will produce a highly skewed distribution, scores will be transformed to ranks prior to statistical analysis.

G. EVALUATION OF COMBINABILITY

After at least 120 patients have completed the study, the possibility of combining these patients with a concurrent European study (Protocol PTL-0022) will be considered. The European study is expected to have enrolled approximately 80 patients by this time. Combinability will be based on three factors.

1. There should be no significant interaction between location (US versus Europe) and treatment efficacy.

2. The US and European population should be similar on demographic and pre-treatment variables and the level of medical care. Variables examined will include:
 - Age
 - Race
 - Body weight
 - Baseline adhesion score
 - Previous and concomitant medications (AHFS classification²)
 - Presence of endometriosis
 - Surgical procedures performed (CPT classification)
 - Estimated blood loss
 - Operative time
 - Baseline clinical laboratory values
 - Length of hospital stay (expected to be longer in Europe)
 - Time to second-look laparoscopy
 - Number of patients lost to follow-up (by reason for discontinuation)

Continuous variables will be analyzed using factorial analysis of variance where one factor is treatment group (LUBRICOAT Gel versus Lactated Ringer's Solution) and the other factor is location (US versus Europe). All statistically significant effects involving location will be considered as possible sources of non-homogeneity that might preclude combination of the US and European data. Categorical variables will be analyzed using categorical models equivalent to analysis of variance with factors for treatment group, location and the interaction between treatment group and location.

3. The US and European control groups should be similar on second-look adhesion scores. This variable can serve as a proxy for subtle differences in medical treatment. The 95% confidence intervals of the difference between the US and European control groups will be presented.

For each of these factors, data will also be analyzed and presented by individual center within the US and Europe.

² McEvoy, G. K., Ed. American Hospital Formulary Service Drug Information. American Society of Health-System Pharmacists, Inc., Bethesda, MD, 1995.

These data will be presented to FDA and the US and European data will not be combined unless Lifecore, Inc. and FDA agree that there are no clinically significant differences that preclude that combination.

If the US and European centers are combinable, then the study will terminate as soon as that decision is made. All patients currently enrolled in the study will be followed to second-look laparoscopy and added to the database for the final statistical analysis.

If the US and European centers are not combinable, then enrollment in the US protocol will continue until 200 evaluable patients have completed the study.

The decision to stop or continue the study will not be affected by the p values of the difference between LUBRICOAT Gel and Lactated Ringer's Solution. Note that:

1. If the data are not combinable, the U.S. study will not be stopped regardless of the statistical significance of the difference between the treated and control groups (either in the combined US/European study or the US study alone).
2. If the data are combinable, and the difference between the treated and control groups in the combined US/European study is not statistically significant, the study will not be continued, but will be stopped and considered to have failed.

Therefore, p values required to demonstrate statistical significance will not be adjusted.

..CONFIDENTIAL..

..CONFIDENTIAL..

H. POWER

Power calculations were performed using the method described by Lachin³ using an alpha level of 0.05 and a beta level of 0.20 (80% power). Preliminary analysis of a Phase I study indicated a mean adhesion score of 1.7 (Standard deviation: 1.4) for the treated group and 5.7 (Standard deviation: 2.7) for the control group. Assuming that 20% of the treatment group and 10% of the control group are lost to follow-up, scoring these patients as treatment failures would yield a mean adhesion score of 4.6 (Standard deviation: 5.9) for the treated group and 6.7 (Standard deviation: 4.1) for the control group. Assuming a standard deviation of 5.0, 180 patients would be required. Thus the 200 evaluable patients (approximately 250 total patients) appears to provide sufficient power to reject the null hypothesis if the observed trends are maintained.

XI. PATIENT DROPOUT RATIONALE

Study enrollment has been planned to allow for a worse case 30% screen failure rate and 20% loss to follow-up rate. This correlates with our request for 350 patients to be asked to participate in the study, with 250 expected to receive treatment, and 200 to complete second-look laparoscopy. All patients assigned study numbers and receiving treatment will be carefully followed and all screen failure and loss to follow-up patients documented. All efforts will be made to keep these to a minimum.

Any patient who fails to return for the Day 7 - 28 laboratory determination and/or the second-look laparoscopy will be contacted and interviewed if possible as to her reason for not returning and her medical status ascertained relative to the effects of the study device. All attempts to contact the patient will be documented on case report form **FINAL STATUS**.

A patient may be discontinued from the study at any time in the event of a serious or intolerable adverse event,¹ the need for an excluded medication, an intercurrent illness, a protocol violation or at the patient's request.

³ Lachin JM. Introduction to sample size determination and power analysis for clinical trials. Controlled Clinical Trials 2:93-113. 1981.

ATTACHMENT B

PILOT STUDY EFFICACY DATA

(from IDE G950025/S9, Final Clinical Study Report, Attachment 10)

NUMBER AND PROPORTIONS OF ADHESIONS

	LUBRICOAT Gel (N=11)		CONTROL (N=9)		P-value
	Mean	SD	Mean	SD	
BASELINE					
Number of Sites with Adhesions	3.55	4.52	4.33	3.93	0.687
Number of Sites where Adhesions were Lysed	2.82	4.04	3.78	3.49	0.582
Number of Primary Surgical Sites	5.36	3.50	5.56	2.50	0.892
INCIDENCE OF ADHESIONS AT "SECOND - LOOK"					
Total Number of Sites with Adhesions	6.09	4.59	11.00	3.24	0.015
				1.34	0.706
					0.023
Total Number of Sites Possible	17.18	1.66	17.44		
Proportion	0.364	0.280	0.629	0.168	

*Students t-test

ATTACHMENT C

TABLES PRESENTED BY LIFECORE AT THE

JANUARY, 2000

GENERAL AND PLASTIC SURGERY

ADVISORY PANEL MEETING

**INTERGEL® Adhesion Prevention Solution
General and Plastic Surgery Panel
Lifecore Biomedical, Inc., (Sponsor)**

TOPIC

PRESENTER

INTRODUCTION

**Georgiann Keyport, M.S., RAC.
Regulatory and Clinical Affairs Manag
Lifecore Biomedical, Inc.**

**CLINICAL STUDY RESULTS
AND CONCLUSIONS**

**Douglas Johns, Ph.D.
Consulting Scientist
Ethicon, Inc.**

AND

**Gere diZerega, M.D.
Consulting Physician
Professor of Obstetrics and
University of Southern California**

Binary Shift Table for AFS Score

INTERGEL

Category	Baseline		Second-Look	
	AFS		Minimal & Mild 0-5	Moderate & Severe 6-10
Minimal & Mild	0-10	122	119	3
Moderate & Severe	11-32	9	9	0
Total		131	128	3

Control

Category	Baseline		Second-Look	
	AFS		Minimal & Mild 0-5	Moderate & Severe 6-10
Minimal & Mild	0-10	117	107	10
Moderate & Severe	11-32	17	10	7
Total		134	117	17

p = 0.003