FDA Executive Summary

Prepared for the
March 20, 2013 meeting of the
Circulatory System Devices Panel

P100009

Abbott Vascular MitraClip Clip Delivery System

INTRODUCTION

This is the FDA Executive Summary for a first-of-a-kind transcatheter mitral valve repair system, the Abbott Vascular MitraClip Clip Delivery System (MitraClip CDS) and accessories. This device has been reviewed by the Division of Cardiovascular Devices within the Center for Devices and Radiological Health of the Food and Drug Administration under Premarket Approval (PMA) application P100009, which is the subject of this Advisory Panel meeting.

This memorandum will summarize the FDA's review of the PMA up to this point, highlighting particular areas for which we are seeking your expertise and input. These topics will include the proposed indications for use, the results of the randomized clinical study, the additional analyses performed to support the indication for use and the proposed post-approval study. At the conclusion of your review and discussion of the data presented, the Agency will ask for your recommendation regarding whether or not the data demonstrate a reasonable assurance of safety and effectiveness.

Table of Contents

1 DEVICE DESCRIPTION

The MitraClip Clip Delivery System (MitraClip CDS) consists of three major components: the MitraClip device, the Delivery Catheter and the Steerable Sleeve. The implant procedure is performed using a steerable guide catheter cleared via the 510(k) regulatory pathway under submissions K112239, K100789, K093866, K091596 and K083793.

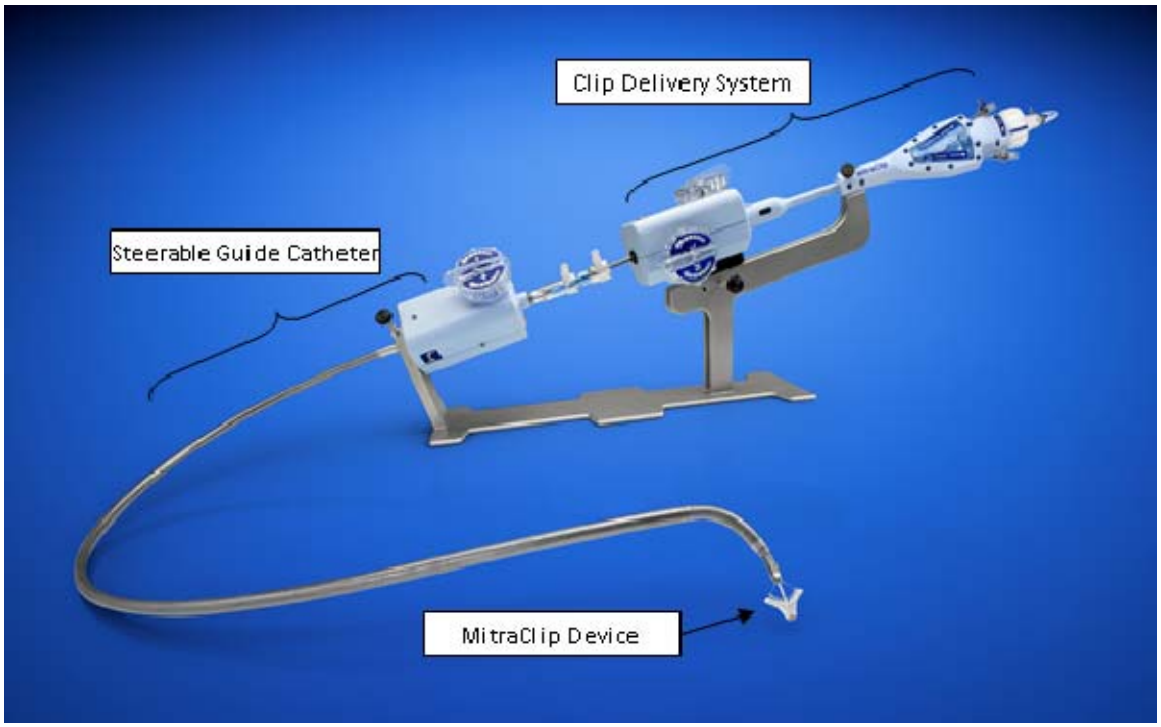


Figure courtesy of Abbott Vascular

The MitraClip device is a single sized, percutaneously implanted mechanical clip for the reduction of mitral regurgitation. The MitraClip device grasps and coapts the mitral valve leaflets resulting in fixed approximation of the mitral leaflets throughout the cardiac cycle. The MitraClip is placed without the need for arresting the heart or cardiopulmonary bypass under general anesthesia. The implantable MitraClip device is fabricated with metal alloys and polyester fabric (Clip cover) that are commonly used in cardiovascular implants.

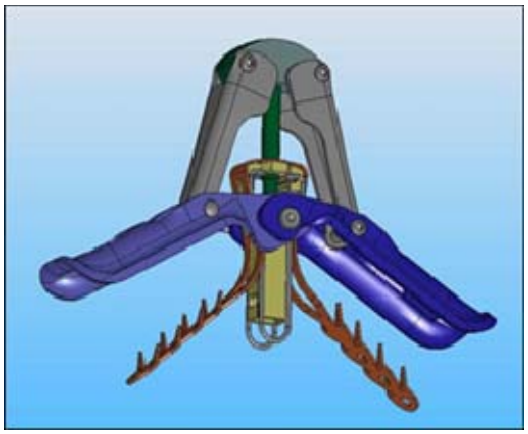


Figure courtesy of Abbott Vascular

The Delivery Catheter is a 10 Fr, long, flexible hydrophilic-coated multi-lumen shaft secured to the MitraClip Device at the distal end and to a handle at its proximal end. The distal tip of the Delivery Catheter is radiopaque to allow visualization under fluoroscopy and is designed to be securely attached to the MitraClip Device. The 24 Fr Steerable Guide Catheter gains access to the left atrium via a trans-septal puncture and positions and orients the Clip Delivery System and MitraClip Device in the appropriate location above the mitral valve.

The Delivery Catheter handle and Steerable Sleeve of the Clip Delivery System position, actuate, and deploy the MitraClip Device.

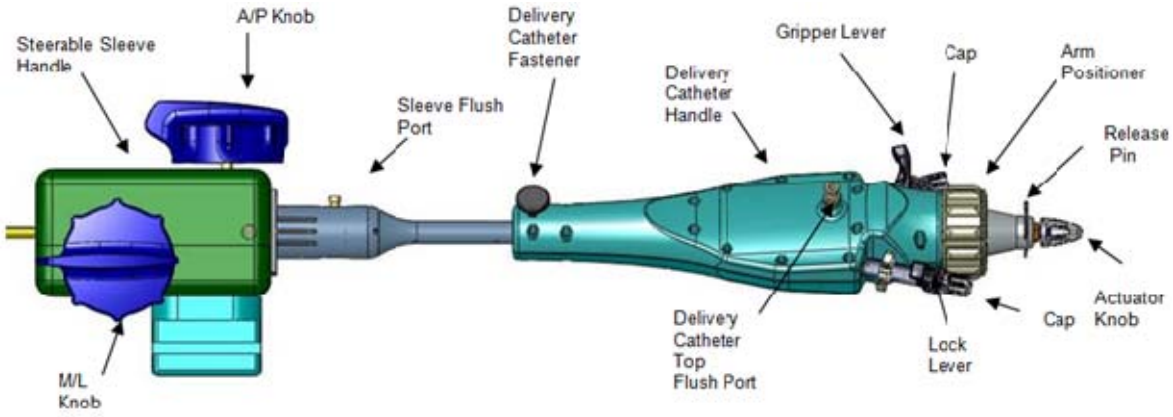


Figure courtesy of Abbott Vascular

The MitraClip CDS enables placement of the MitraClip device on the mitral valve leaflets resulting in permanent leaflet approximation and the formation of a double orifice (pictured on the right below), based on a surgical repair technique sometimes referred to as the Alfieri procedure/technique (pictured on the left below).

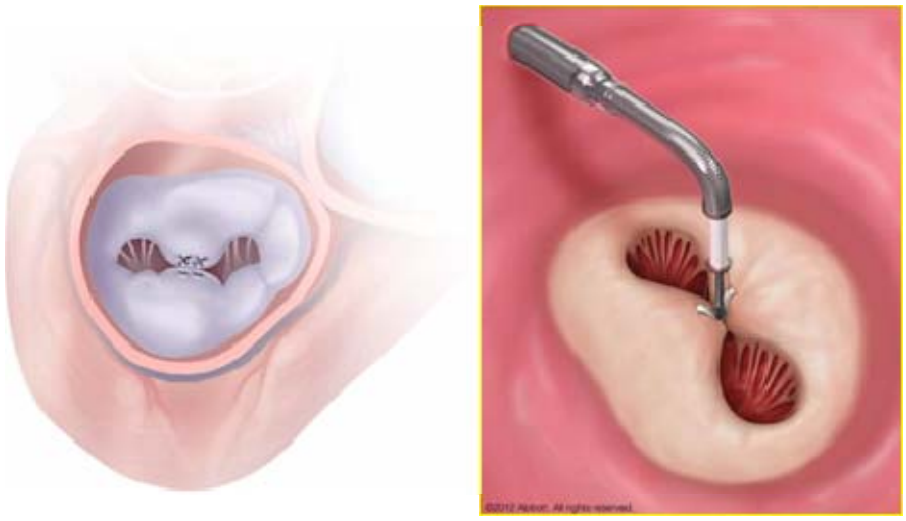


Image: Feldman et al. J Am Coll Cardiol. (2000).

This surgical technique was rarely used in the United States by experienced mitral valve surgeons but when performed was usually accompanied by annuloplasty ring placement The Alfieri surgical technique is rarely used as a stand alone procedure in the current era because of unsatisfactory results when performed without a concomitant mitral annuloplasty.

2 PROPOSED INDICATIONS FOR USE

Abbott Vascular originally proposed the following Indication for Use for the MitraClip Delivery System, to be primarily supported by the EVEREST II RCT:

The MitraClip Delivery System is intended for the reconstruction of mitral valve insufficiency through tissue approximation in patients with significant mitral regurgitation.

After hearing some of FDA's concerns about the ability of the EVEREST II RCT data to support a PMA approval, the sponsor performed additional *post-hoc* analyses on other patient groupings and changed the originally proposed Indication for Use to the current version, which we will be considering today:

The MitraClip Clip Delivery System is indicated for the percutaneous reduction of significant symptomatic mitral regurgitation (MR≥3+) in patients who have been determined by a cardiac surgeon to be too high risk for open mitral valve surgery and in whom existing co-morbidities would not preclude the expected benefit from correction of the mitral regurgitation.

FDA Comment: The currently proposed indication for use was developed by Abbott Vascular after study results were known and analyzed by Abbott Vascular (*post-hoc*) and was advanced as an alternative after FDA communicated their concerns that the evidence

available to support a finding of safety and effectiveness of the device for the originally proposed indication for use and population for use was insufficient. FDA believes the evidence necessary for determination of safety and effectiveness sufficient for approval of a first of a kind device should not be based on a retrospective evaluation of registry data re-configured to support an indication for use and population for use developed *post-hoc*.

3 PRE-CLINICAL STUDIES

3.1 *In Vitro* Testing

The sponsor conducted *in vitro* performance and characterization studies of the MitraClip Delivery System:

- Test results demonstrated that the device is compliant with FDA recognized international standards for biocompatibility.
- Packaging and sterilization processes were validated according to FDA recognized international standards as well.
- The MitraClip was evaluated for MRI compatibility.
- FDA performed a comprehensive review of the pre-clinical bench testing performed under challenging conditions to verify the design of the MitraClip Delivery System.
  - The testing was conducted in accordance with the international heart valve standard (ISO 5840) and the draft FDA heart valve guidance document.
  - Testing included fatigue (15 years of simulated use) and corrosion evaluation of the MitraClip Delivery System.
  - The results of the bench testing performed supported device safety in the anticipated clinical environment for the intended patient population.
- Design verification testing of the accessories was done and was found to be acceptable.

FDA Comment: FDA has no remaining concerns regarding the pre-clinical bench testing.

3.2 *In Vivo* Animal Testing

The sponsor conducted several *in vivo* performance and characterization studies of the MitraClip delivery system. There were 37 animals utilized across three GLP and two non-GLP studies. Of the 37 animals, 28 were utilized in GLP work using the final design iteration across three of the five referenced protocols. The most important study for this device is the chronic evaluation of safety which was performed in 21 of the 37 animals. All of the animals in the chronic study received their device via open visualization through left thoracotomy. There were two cases of symptomatic endocarditis (10%), and four more asymptomatic cases detected at necropsy (19%) for a total infectious and/or inflammatory endocarditis rate of 28%. The sponsor explained that these events were likely due to surgical contamination.

The second GLP study was acute and studied three animals where placement occurred via open visualization through left thoracotomy while the last GLP study was acute and studied four animals in which the device was received via femoral vein access and subsequent atrial septal puncture to access the mitral valve.

FDA Comment: FDA has no concerns regarding the pre-clinical animal testing.

4 DEVICE MODIFICATIONS

During the pivotal trial, the design of the MitraClip Delivery System continually evolved but FDA believes that the changes made were minor. During the REALISM CAP, Abbott Vascular initiated a voluntary suspension of enrollment due to several occurrences of radio-opaque ring detachment. Abbott performed a root cause analysis of this device malfunction and made a modification to the device to correct the cause. No further incidents of radio-opaque ring detachment have been discovered.

FDA Comment: FDA has no concerns with the device modifications and believes that the clinical data collected in the EVEREST II trial are applicable to the current design of the device and delivery system proposed in this PMA application.

5 REGULATORY HISTORY

The MitraClip Clip Delivery System (MitraClip CDS or MitraClip) was originally manufactured by Evalve, Inc. Abbott Vascular acquired Evalve, Inc., in September 2009. The EVEREST I feasibility study was approved with conditions on April 16, 2003. The EVEREST I study was intended to evaluate safety at 30 days and establish feasibility of repair using the MitraClip CDS.

FDA conditionally approved the EVEREST II pivotal clinical trial on November 3, 2004. The EVEREST II RCT was designed as a prospective, randomized, active controlled, multi-center clinical trial to evaluate the safety and effectiveness of the MitraClip in the treatment of moderate-to-severe (3+) or severe (4+) chronic mitral regurgitation (MR). The RCT was designed to demonstrate superiority of the device to mitral valve surgery for the primary safety endpoint and to demonstrate that the device would be no worse than mitral valve surgery for the primary effectiveness endpoint. Beginning prior to initial study approval, FDA repeatedly expressed many concerns with the trial design, including, but not limited to the use of MR $\leq$ 2+ as the criterion for success in the primary effectiveness endpoint, a large margin of reduced effectiveness for the primary endpoint hypothesis, the heterogeneity of MR etiology, and the relative inexperience with mitral valve repair of a substantial proportion of the surgeons operating on the control group. A single arm registry was conditionally approved for "high risk" non-operative patients (High Risk Registry or HRR) on November 16, 2006. FDA repeatedly expressed the concern that this registry study, without a rigorous pivotal trial with positive results, could not by itself support PMA

approval of the MitraClip CDS due to a number of study limitations (described in more detail below).

The REALISM continued access protocol (CAP) was approved with conditions in November 2008. This trial allows enrollment for patients based on the same inclusion and exclusion criteria as those in EVEREST II. 545 patients have been treated and completed 1 year of follow-up as part of the REALISM study reported in the PMA. This continued access protocol includes an extension of the High Risk Registry – REALISM High Risk.

Although the PMA originally proposed approval for a different indication for use based on the RCT results, this document focuses primarily on the *post-hoc* developed indication for use proposed in section 2.0 above and the retrospective subset analyses submitted to FDA in support of this indication for use.

The FDA has been working with the Sponsor on a well-designed new trial to establish the safety and effectiveness of the MitraClip device in a well-defined population. In February 2012, FDA conditionally approved the COAPT clinical trial which is a prospective, randomized, active controlled, multi-center clinical trial to evaluate the safety and effectiveness of the MitraClip device in the treatment of symptomatic (NYHA Functional Class II, III or ambulatory IV) functional mitral regurgitation (≥3+) in patients that have co-morbidities that preclude surgery (i.e., the probability of death or serious morbidity exceeds the probability of meaningful improvement). For more information on the COAPT study, please visit: http://www.clinicaltrials.gov/ct2/show/NCT01626079?term=mitraclip&rank=3.

FDA Comment: FDA believes that the COAPT trial is a reasonable pivotal trial and that the prospective analyses of this trial may support PMA approval of the MitraClip CDS for the specific subset of extremely high surgical risk (too high risk for surgery) heart failure patients with functional mitral regurgitation.

6 SUMMARY OF KEY FINDINGS

As documented throughout the remainder of this memorandum, FDA's review of the data included in this PMA has raised a number of important concerns, which can be summarized as follows:

- The EVEREST II RCT did not demonstrate an appropriate benefit-risk profile when compared to standard mitral valve surgery in a selected mitral valve patient population.
- For a variety of reasons, the EVEREST II HRR single arm registry data are not easily interpretable.
- REALISM HR is a continued access protocol cohort that was not intended to be used as a pivotal data set and is difficult to interpret.

- The Integrated High Surgical Risk Cohort, developed by pooling two registry data sets in a *post-hoc* manner, has major design limitations.

- The Duke Propensity Score Analysis was a retrospective, subset analysis with results that are difficult to interpret and where the matched cohorts do not represent any well-defined population.

In the following Clinical Studies section, we will present more detail regarding our review of the data, as well as expand on these concerns.

7 CLINICAL STUDIES

7.1 EVEREST I

The EVEREST I feasibility study enrolled 55 patients in a single-arm registry to refine the protocol prior to commencing the pivotal EVEREST II study. The FDA and the Sponsor agreed that these results would not be pooled with the EVEREST II data.

FDA Comment: EVEREST I data (feasibility study) combined with pre-clinical bench and animal testing supported the conclusion that sufficient safety existed to proceed with a pivotal trial.

7.2 EVEREST II

7.2.1 Randomized Controlled Trial

7.2.1.1 Study Design

The EVEREST II clinical trial was a 279-patient randomized study with open surgical mitral valve operation as the control. This study enrolled mitral regurgitation patients with grade 3+ to 4+ mitral regurgitation who were randomized to either the MitraClip or mitral valve surgery. Patients were randomized 2:1 to the Device group (184) and the Control group (95), of which 21 patients (6 Device and 15 Control) did not undergo treatment per their randomized assignment, and are reported as "randomized not treated" (RNT).The proposed Indications for Use and the inclusion criteria for the study included a wide spectrum of patients that were anatomically and clinically amenable to either MitraClip or surgical valve repair procedures. Patients with functional (ischemic and non-ischemic), and all degenerative etiologies (anterior and/or posterior leaflet and chordae) were included. Functional MR patients comprised 27% of patients enrolled.

FDA believes that combining all etiologies of mitral regurgitation in this moderately sized study is problematic. FDA indicated to the sponsor prior to the onset of the trial that the study should be powered separately for functional (ischemic and non-ischemic patients) and degenerative etiologies (spectrum of limited prolapse to Barlow's).

7.2.1.2 Study Results

7.2.1.2.1 Safety

The primary safety endpoint was a non-hierarchal composite of 12 major adverse events (MAE) at 30 days (including death, MI, reoperation for failed repair [surgical group only], non-elective cardiovascular surgery for adverse events, transfusion of 2 or more units of blood or blood products, etc.). The superiority hypothesis was that the proportion of MitraClip patients who had a MAE would be at least 6% less than the proportion of surgical control patients who had a MAE using the Per Protocol analysis set. This objective was met for the composite safety endpoint; however, when examining these results closer, it was noted that the difference between the two groups was driven by the transfusion (>2 units) component of the composite. Excluding transfusion, the per protocol combined procedural and 30-day MAE rate was 0.7% (1/137) for MitraClip patients and 11.3% (9/80) for surgery patients. When considering the all treated population, these rates become 4.5% (8/178) in the MitraClip group and 11.3% (9/80) for the surgery patients. There were no differences in death, strokes, myocardial infarction or infection. The transfusion rate in the surgery group was 47.5%. One reason for this relatively high rate of transfusion is that many of the surgery patients who had bleeding also had concomitant operations (MAZE, multiple valve replacement/repair, coronary bypass, reduction in size of LA, ligation of LAA, etc.) which could affect the complication rate. The performance of concomitant procedures in close to 50% of surgical patients also makes strict comparison to the device group challenging, considering that none of the device patients underwent concomitant procedures at the time of MitraClip placement.

Table 1: Adjunctive Surgery All Treated Patients – Control Group (N = 80)

Adjunctive Surgery Type[a]	(%) n/N
Major Procedures	**16.2% (13/80)**
Other Valve Surgery	5.0% (4/80)
CABG	10.0% (8/80)
CABG + Valve Surgery	1.2% (1/80)
Other Procedures[b]	**31.3% (25/80)**
None	**52.5% (42/80)**

[a] *Adjunctive surgery as described in the operative report.*

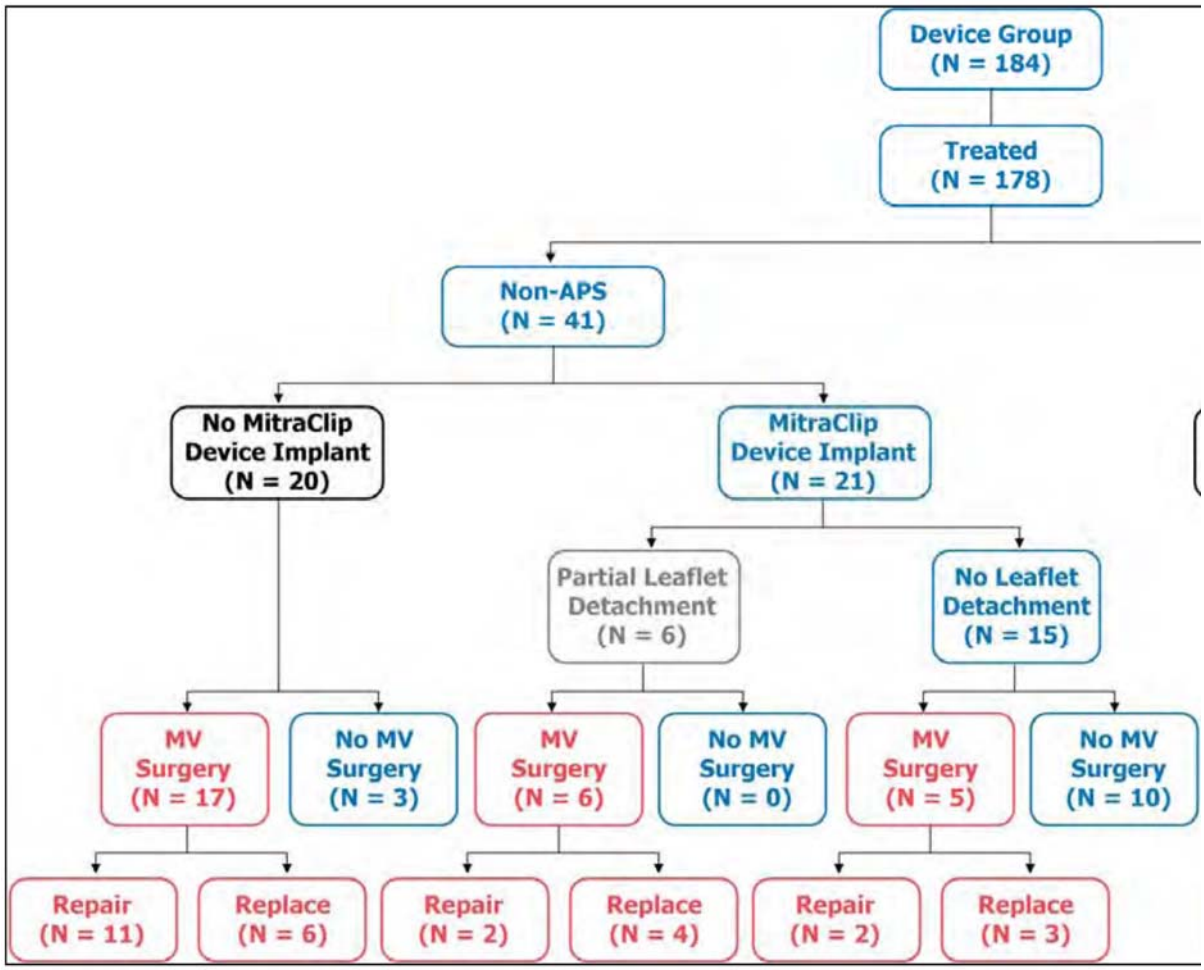
[b] *Other Procedures include MAZE, PFO Closure, ASD Closure, Atrial Appendage Closure or Removal; patients may have had more than one of these procedures; patients included under Major Procedures are not included under Other Procedures.*

Importantly, a high proportion of the patients with complications counting towards the safety composite were operated on by mitral valve surgeons with fewer than 25 total operations (repair + replacement) in the previous year, suggesting procedural experience and expertise may have been skewed against surgical control treated patients. As an example, 37% of patients with the Major Adverse Event of bleeding were operated on by very low volume mitral repair surgeons (<15 mitral repairs/year).

Sixty-nine of 80 treated control patients (86.3%) had their valve successfully repaired. Successful repair was predicted in 75 control patients (94% of all patients) prior to surgery and was achieved in 92% (69/75) of those in whom it was predicted. Five control patients

underwent valve replacement as was predicted pre-operatively, as did the 6 patients in whom predicted MV Repair could not be achieved. One hundred seventy-eight (178) of 184 patients underwent a procedure with the intent to perform a MitraClip implant(s). Acute Procedural Success (APS), defined by the sponsor as successful implantation of the MitraClip device with resulting MR ≤ 2+ at discharge was achieved in 77% (137/178) of patients in whom it was attempted (i.e., 23% acute procedural failure rate, 41/178). Early failure due to single leaflet device attachment (SLDA) occurred in eight of the 158 (5.1%) patients implanted with a MitraClip device and resulted in the need for surgery; two additional failures due to SLDA occurred at 9 months and at 1 year. Of these, surgical repair was performed in 50% with replacement in the remainder. A full accounting of outcomes in patents where APS was not achieved is presented below in Figure 1.

Figure 1: Treatment Course for Patients in the Device Group without APS (as defined by sponsor)



FDA defined APS as successful implantation of the MitraClip device with resulting MR ≤ 1+ (instead of MR ≤ 2+ for Sponsor's definition) at discharge. Based on FDA's definition, 46.6% (83/178) of device treated patients achieved APS (with MR ≤ 1+).

FDA indicated to the Sponsor that a 30-day primary safety endpoint was not acceptable as the only safety endpoint. Safety needed to be evaluated over the course of the study.

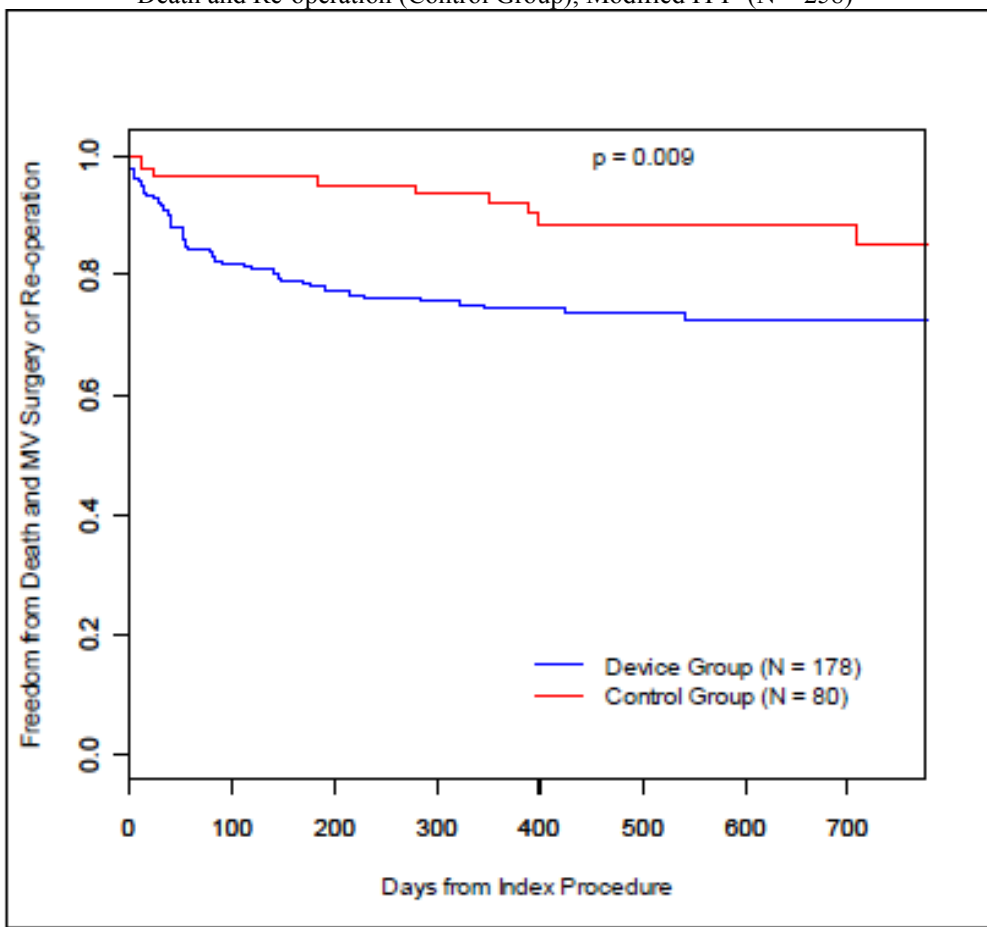
Excluding major bleeding, the sponsor provided the major adverse events (MAE) complication profile for each group over 12 months as follows:

Table 2: MAE rate for Device and Control patients excluding Transfusion

	Procedural	Procedure to 30 days	30 days to 12 months
Device	1.1%	2.8%	9.6%
Surgery Control	8.8%	5.0%	7.5%

In addition, for the 2 most important parameters of death and need for subsequent mitral valve surgery (device group) or death and re-operation (surgery control group), the following results were demonstrated:

Figure 2: Kaplan-Meier Freedom from Death and Mitral Valve Surgery (Device Group) and Freedom from Death and Re-operation (Control Group), Modified ITT[a] (N = 258)



[a] *Modified ITT includes all patients randomized to either device or control minus those randomized and not treated for each group*

After careful evaluation of each component of the composite, FDA does not believe that this percutaneous device has a superior safety profile over time to that of open operation in the population studied.

7.2.1.2.2 Effectiveness

The primary effectiveness endpoint proposed by sponsor was Clinical Success at 12 months, defined as freedom from death, MR>2+, and mitral valve reintervention at 12 months. The primary effectiveness hypothesis proposed by the sponsor was to demonstrate that the clinical success rate at 12 months in the device group was no more than 31% worse than the surgical control group (i.e., margin of 31%).

During study design, the FDA disagreed with the definition of the primary effectiveness endpoint proposed by the sponsor and recommended that freedom from MR>2+ should be replaced by freedom from moderate (2+) or higher MR (i.e., MR >1+). FDA also notified the sponsor that the proposed margin of reduced effectiveness of 31% was quite large for a population expected to be excellent candidates for MV Repair based on anatomic inclusion and exclusion criteria.

Despite the concerns regarding the Sponsor's Primary Safety and Effectiveness Endpoints, the criteria for success chosen by the Sponsor were met (Table 3. The Sponsor's Per Protocol analysis for effectiveness includes:

- For the treated device group
  - 137 device treated patients where acute procedural success was achieved are included, and
  - 41 patients in whom acute procedural success using the device was not achieved were excluded.
- For the treated control group
  - All treated patients randomized to surgery were included (n=80)

Table 3: Sponsor Defined Safety and Effectiveness Endpoint Summary, EVEREST II RCT, Based on Sponsor Defined Per Protocol cohort[a]

Analysis Cohort [d]	**Safety Superiority [b]**	**Effectiveness Non-Inferiority [c]**
	Per Protocol (N = 217)	**Per Protocol (N = 217)**

Delta	-6%	-31%
Endpoint Rate (Device Group)	9.6% (13/136)	72.4% (97/134)
Endpoint Rate (Control Group)	57.0% (45/79)	87.8% (65/74)
Difference	-47.4%	-15.4%
p-value	< 0.0001	0.0012
97.5% UCB (for safety) or 95% LCB (for effectiveness) Device-Control	-34.4%	-25.4%
Endpoint Met/Not Met	**Met**	**Met**

[a] *Per Protocol analysis includes patients in the Device group who achieved APS and patients in the Control group who underwent surgery.*
[b] *Safety endpoint is based on MAE rate at 30 days*
[c] *Effectiveness endpoint is Clinical Success defined as freedom from death, surgery after APS or index surgery, and MR > 2+.*
[d] *Sample sizes less than that shown in the header reflect missing data.*

Based on FDA defined primary effectiveness endpoint (Freedom from death, reintervention and MR > 1+) for the FDA defined per-protocol cohort, the 95% lower confidence bound for difference in 12 months clinical success rate between device and control group was -37.7% (p=0.17).

Table 4: Analysis using FDA Primary Effectiveness Endpoint - Freedom from death, mitral valve surgery, and MR > 1+ at 12 months

Analysis Cohort	**Effectiveness (margin of reduced effectiveness)**
	FDA Defined Per Protocol (N=163)
Delta	-31%
Endpoint Rate (Device Group)	45.1% (37/82)
Endpoint Rate (Control Group)	68.9% (51/74)
Difference	-23.8%
p-value	0.1692
95% LCB Device – Control	-37.7%

Of the 80 patients treated with surgery in the Control group, MR severity ratings are missing for 3 patients due to missing/non-evaluable echocardiograms and death prior to discharge.

All of the remaining 77 patients had discharge MR severity rating of ≤ 2+ and 84.4% (65/77) achieved an MR severity rating of ≤ 1+ at discharge.

Table 5: MR Severity at Discharge,[a] All Treated Patients (N=258)

MR Severity	Device Group (N = 178) % (n/N)	Control Group (N = 80) % (n/N)
0: None	1.2% (2/173)	22.1% (17/77)
1+: Mild	51.4% (89/173)	62.3% (48/77)
1+ to 2+: Mild-to-Moderate	15.0% (26/173)	5.2% (4/77)
2+: Moderate	16.2% (28/173)	10.4% (8/77)
3+: Moderate-to-Severe	11.0% (19/173)	0.0% (0/77)
4+: Severe	5.2% (9/173)	0.0% (0/77)
Death prior to discharge	1	1
Discharge and 30 day Echo not available/evaluable	4	2

[a] *30-day MR severity is used if discharge MR severity is missing*

The MR severity seen on echocardiography at 12 months follow-up was different for device versus control treated patients as seen in Table 6 below.

Table 6: MR Severity at 12 Months (Reproduced from EVEREST II RCT Report)

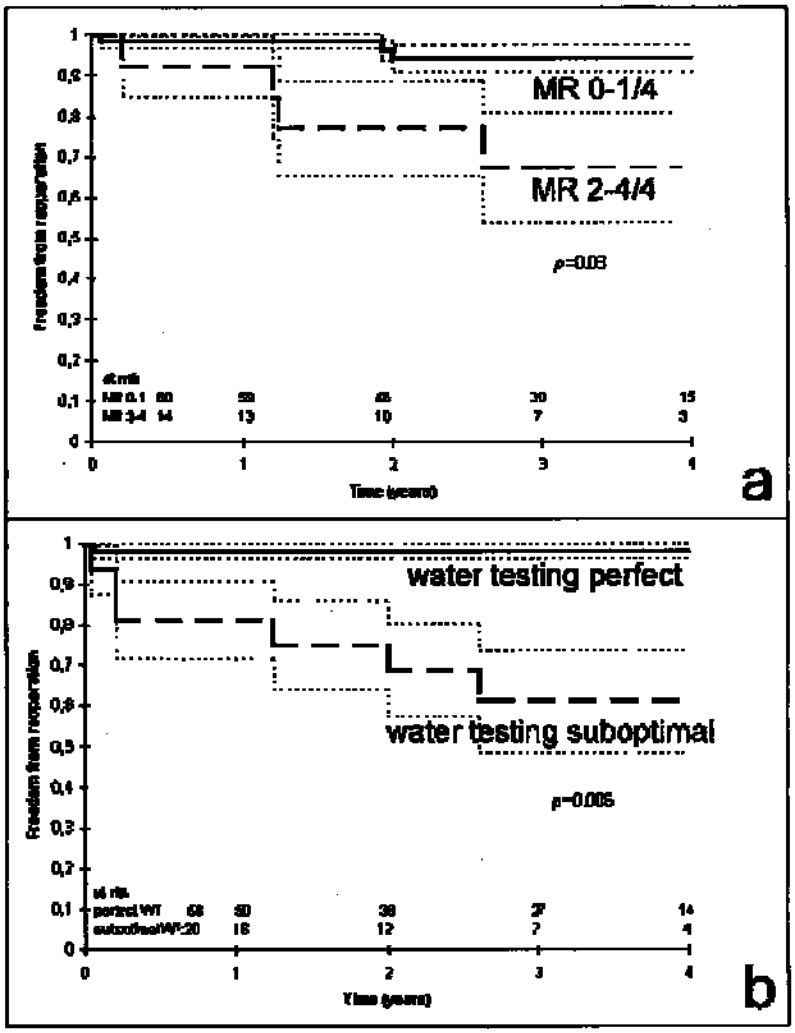
MR Severity at 12 Months[a]	Device Group (N = 184) % (n/N)	Control Group (N = 95) % (n/N)
0: None	3.4% (5/149)	17.9% (12/67)
1+: Mild	36.2% (54/149)	58.2% (39/67)
1+ to 2+: Mild-to-Moderate	12.1% (18/149)	7.5% (5/67)
2+: Moderate	27.5% (41/149)	13.4% (9/67)
3+: Moderate-to-Severe[b]	16.1% (24/149)	3.0% (2/67)
4+: Severe	4.7% (7/149)	0.0% (0/67)
Death	11	5
Surgery in APS Device Group/Re-op in Control Group	9	2
RNT	6	15
MR Not Evaluable/Echo not done/Withdrawn	9	6

[a] *18-month or 24-month MR is used for 4 patients ([redacted]) in the Device group and 2 patients ([redacted]) in the Control group. MR severity is not presented for RNT patients and patients who failed the Clinical Success endpoint due to death, and surgery (for APS Device group) or re-operation (Control group).*

[b] *3 patients in the Device group ([redacted]) who did not achieve APS and withdrew without undergoing surgery are included with MR of 3+ at 12 months.*

Per FDA request, the Sponsor also provided 24-month follow-up data for the Primary Effectiveness endpoint using FDA recommended definition of freedom from death, reintervention or surgery, and >1+ MR. FDA supports this more rigorous threshold based on data from Alfieri's own mid-term follow-up of patients undergoing the "Alfieri procedure" which forms the theoretical basis for performance of the MitraClip procedure. Mid-term follow up in these patients show worse long-term results when more than mild (>1+) residual MR is seen on the post-repair echocardiogram and water testing of the operative repair reveals residual MR (less than perfect). The degree of water-testing leakage was linearly correlated to TEE MR grading (r^2 = 0.589, p = .0001). Abnormal water-testing results predicted a post-operative MR grade > 1 with 93% sensitivity, 89% specificity, 65% positive predictive value, and 98% negative predictive value. The rehospitalization rate was higher in those patients with suboptimal reconstruction: freedom from rehospitalization at 4 years was 67% ±33% in those patients with normal competence at intraoperative water testing, whereas it was 56% ± 43% in those who had some degree of leak at intraoperative testing ($P < 0.01$).

Figure 3: Freedom from Reoperation based on residual MR post repair: Midterm follow-up of patients undergoing the Alfieri procedure (Massiano et al. J Thorac Cardiovasc Surg 2003;126: 1987-97)



In this paper, Alfieri concludes:

> "*Therefore, particularly when annuloplasty is not added to the repair procedure, only optimal competence of the valve should be considered acceptable to avoid the risk of late progression of the disease with recurrence of severe regurgitation and need for reoperation.*"

FDA believes the appropriate population for analysis regarding effectiveness is the Modified ITT population. Modified ITT (MitraClip n=178, control n=80) includes all randomized patients who had an attempted procedure. This includes patients who did not achieve acute procedural success and excludes Randomized Not Treated (RNT) patients from each arm of the study.

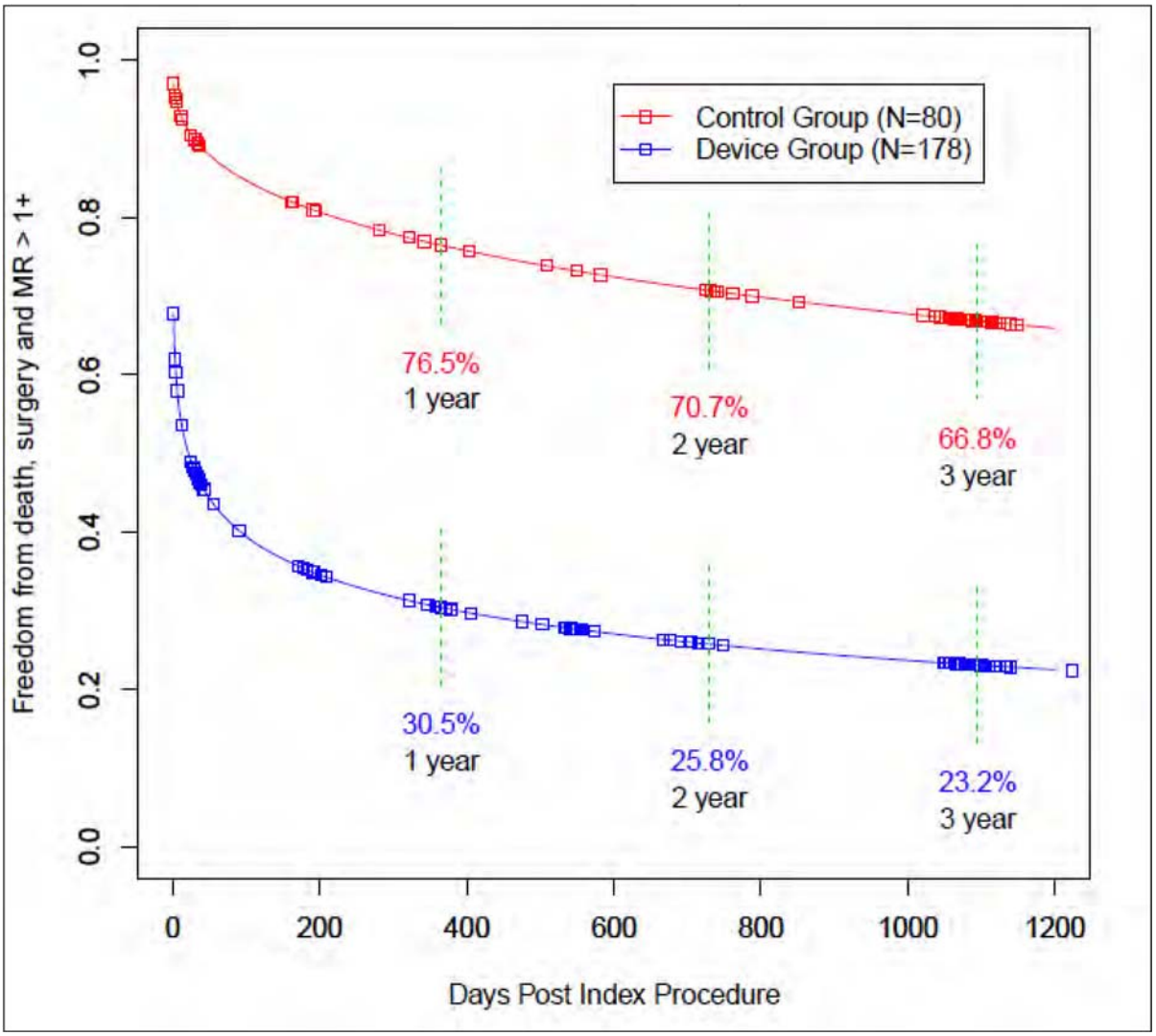
Table 7: Components of Failure of Clinical Success (MR ≤ 1+) at 12 and 24 Months, All treated patients (n= 258)

Component of Failure	12 Months			24 Months		
	Device Group (N = 178) % (n/N)	Control Group (N = 80) % (n/N)	Device - Control (95% two-sided CI)[a]	Device Group (N = 178) % (n/N)	Control Group (N = 80) % (n/N)	Device - Control (95% two-sided CI)[a]
Death[b]	4.6% (8/175)	6.8% (5/74)	-2.2% (-9.6%, 5.3%)	8.9% (15/169)	11.4% (8/70)	-2.6% (-12.2%, 7.1%)
MV surgery for MV dysfunction after clip implant (Device group) or Re-operation of the mitral valve for MV dysfunction (Control group)	21.1% (37/175)	2.7% (2/74)	18.4% (10.4%, 26.5%)	22.5% (38/169)	4.3% (3/70)	18.2% (9.3%, 27.1%)
MR > 1+	49.7% (87/175)	21.6% (16/74)	28.1% (15.2%, 41.0%)	52.7% (89/169)	17.1% (12/70)	35.5% (22.9%, 48.1%)
Total	**75.4% (132/175)**	**31.1% (23/74)**	**44.3% (31.1%, 57.6%)**	**84.0% (142/169)**	**32.9% (23/70)**	**51.2% (37.8%, 64.5%)**

a Confidence intervals are calculated based on asymptotic methods for the difference between two independent binomial proportions with continuity correction.
b In the Device group 4 patients ([redacted] underwent MV surgery and subsequently died. In the Control group 1 patient [redacted] underwent re-operation and subsequently died. These patients are included as failure due to MV surgery/re-operation rather than death.

A Weibull plot of the Freedom from Death, Surgery (Device group) or Reoperation (Control group) and MR > 1+ shows similar results in Figure 4 below.

Figure 4: EVEREST II RCT – Weibull Freedom from Death, Surgery (for Device group) or Re-operation (for the Control group) and MR > 1+



A very large difference in the clinical effectiveness endpoint as defined by the FDA criteria was demonstrated. In addition, the lower percentage reduction in MR to 0 or 1+ with the device compared to control was also supported by the lower change in LV function/dimensions in the device group as measured by the echo core lab.

The risk for subsequent Mitral Valve Surgery in the Device group was substantially higher than the risk of reoperation in the Control group early after the respective procedures. After 6 months, the linear rates for the late risks for each equalized, as shown in Figure 5 below.

Figure 5: EVEREST II RCT – Kaplan-Meier Freedom from Mitral Valve Surgery (Device Group) or Re-operation (Control Group) – 6-Month Landmark Analysis All Treated Patients (N = 258)

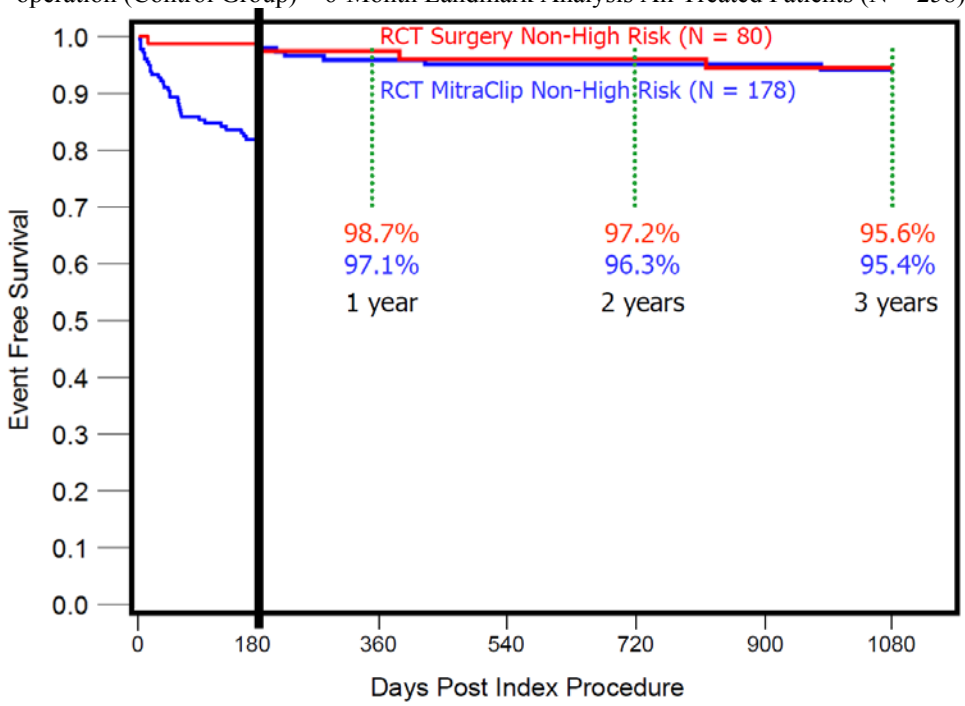


Table 8: Change in LV Measurement from Baseline to 12 Months, Sponsor Defined Per Protocol Cohort (N = 217),
Matched Cases[a]

LV Measurement	Device Group (N = 137)	Control Group (N = 80)	p-value (two-sided)[b]
LVEDV, ml			
Mean ± SD	-21.3 ± 24.1	-40.2 ± 36.2	0.0003
(N)	(118)	(65)	
LVIDd, cm			
Mean ± SD	-0.4 ± 0.5	-0.6 ± 0.6	0.0030
(N)	(122)	(66)	
LVESV, ml			
Mean ± SD	-4.4 ± 14.0	-5.1 ± 20.8	0.7888
(N)	(118)	(65)	
LVIDs, cm			
Mean ± SD	-0.1 ± 0.5	-0.0 ± 0.6	0.4070[c]
(N)	(120)	(66)	

[a] *Only patients who had a measurement at both Baseline and 12 months are included.*
[b] *Two-sided p-value is based on a two-sample t-test. Between group changes are assessed at 5% significance level.*
[c] *No conclusions are drawn from this comparison of LVIDs since this variable did not show statistically significant improvement in the sequential testing within the Device group.*

These measurements were only performed in patients with both baseline and 12-month measurements, hence patients who either died (n=18 Device, n=8 Control) or withdrew (i.e., two groups of patients in whom benefits in terms of LV dimension reduction may not

necessarily be expected) were not included. Also excluded for the device group were the patients treated but without acute procedural success (n=41).

7.2.1.3 Additional FDA Comments on the EVEREST II Trial

A major premise of the EVEREST II study was that there is no downside and the patient can go on to get an operation if the percutaneous procedure is ultimately unsuccessful. This is the basis for the Sponsor's "treatment strategy" argument. However, it is unclear if acute injury occurred to the native valve from placement of, or an attempt to place, the device. In other cases, the valve was injured because of intense fibrosis which affects the ability to remove the clips without injuring the underlying valve or chordae, thus preventing valve repair or making repair much more difficult. Therefore, it is difficult to determine whether the valve was injured during the MitraClip procedure or during the attempt to surgically remove the clips; the total rate of valve injury (including holes in one or both leaflets or torn chordae due to difficulty disengaging the clip or due to fibrosis at the sites of clip attachment) was 26% (10/38) of RCT patients who went to operation and 39% (7/18) of roll-in patients. In addition, acute cardiac or great vessel injury from the procedure led to the need for emergency operations in three patients. For subjects initially treated with the MitraClip who eventually went on to require MV repair or replacement due to residual or recurrent MR, a significant proportion (53%, 20/38) had residual ASDs repaired at the time of surgery. All of these factors increase the complexity and risk of subsequent operation. For device patients who underwent later operation for mitral valve surgery, those with functional etiology required a preponderance of valve replacement procedures (>70% in operative reports submitted). Due to the presence of excess and flexible valvular leaflet tissue in degenerative etiology patients, the opposite surgical procedure pattern was seen (>70% repaired).

It should be noted that half (14/28) of the patients who did not have Acute Procedural Success had mitral valve surgery delayed for more than 30 days (three of these were delayed more than 3 months). In addition, although all of the patients in the study were indicated for operation when enrolled, approximately one third (13/41) of patients who had a failed attempt at device placement did not undergo operation during the study.

Inclusion criteria for the EVEREST II RCT made valve repair rather than replacement highly likely by experienced surgeons; however, a substantial proportion of the patients in both the control group (14%, 11/80) and in the Clip group (45%, 17/38) who received operations underwent valve replacement rather than repair.

The above considerations are quite important from a patient management perspective because acute procedure success was only noted in 77% of MitraClip patients. Hence, a fair number of MitraClip patients would be expected to need surgery following an unsuccessful MitraClip attempt at fixing the mitral regurgitation.

FDA Comment: The EVEREST II RCT did not demonstrate an appropriate benefit-risk profile when compared to standard mitral valve surgery in a selected mitral valve patient population. FDA believes that there are fundamental study design and conduct problems associated with the EVEREST II RCT. Regardless, results demonstrate that acute procedure

success will not be achieved in a significant number of patients despite rigorous anatomic criteria (23%). In addition, using a standard definition of unacceptable MR (>1+), clinical effectiveness at one year is limited compared to the surgical control, and the rate of reoperation in the device group for mitral regurgitation by one year is increased. Except for the difference in transfusion rates between the two groups which may be partially accounted for by a high rate of concomitant procedures in surgical controls, a clinically meaningful safety benefit for device therapy cannot be demonstrated. These results do not indicate that the MitraClip has an acceptable benefit-risk ratio in a MR population that can be treated using either standard cardiac surgery or the MitraClip device.

7.2.2 High Risk Registry (HRR)

The HRR was approved by the FDA as an adjunctive single-arm registry whose data would be evaluated in conjunction with the data from the EVEREST II RCT. The HRR enrolled 78 patients. Patients were considered high surgical risk if either their STS calculated risk was ≥12%, or the surgeon investigator determined the patient to be high risk (≥12% predicted operative mortality) due to the presence of, at a minimum, one of the following pre-specified risk factors used as inclusion criteria:

1. Porcelain aorta or mobile ascending aortic atheroma
2. Post-radiation mediastinum
3. Previous mediastinitis
4. Functional MR with EF < 40%
5. Over 75 years old with EF < 40%
6. Prior re-operation with patent grafts
7. Two or more prior chest surgeries
8. Hepatic cirrhosis
9. Three or more of the following STS high risk factors:
   i. Creatinine > 2.5 mg/dL
   ii. Prior chest surgery
   iii. Age over 75
   iv. EF < 35%

Patients whose STS-score was <12%, but had at least one of the above risk factors were assigned a risk by the surgeon investigator of at least 12% per the protocol

Key to understanding the assignment of high risk (or any predicted surgical risk) is the use of the STS risk calculator. The STS database is the world's largest cardiac surgical database, surpassing 4 million cardiac operations in 2011. A validated mortality risk predictor for MV replacement has been available for online use since 1999. The ability to enter preoperative data specifying the operative procedure as Mitral Valve Repair began in 2001. Collection of these data allowed for development and construction of separate validated models which allow distinct predictions of mortality risk for both MV Repair and MV Replacement. These models were published in 2008, and online calculators for each were available in 2008.

O'Brien SM, Shahian DM, Filardo G, et al, for the Society of Thoracic Surgeons Quality Measurement Task Force. The Society of Thoracic Surgeons 2008 cardiac surgery risk models: part 2—isolated valve surgery. Ann Thorac Surg 2009; 88(Suppl):23– 42.

The enormity of data within the STS database has also allowed development of established risk models for select in-hospital post-operative complications. Specific validated mortality and post-operative morbidity risk algorithms are available for coronary artery bypass procedures, aortic and mitral valve procedures, and their combination. For mitral valve disease, separate validated mortality risk calculations are specifically available for both repair and replacement procedures. Mortality risk calculations for each procedure are referred to as the STS "predicted risk of mortality." In keeping with the Sponsor's preference we will use the terminology of "STS score" when referring to the objectively calculated STS predicted risk of mortality. FDA acknowledges the additional risk calculator for MV Repair was not available to the Sponsor throughout the bulk of the HRR enrollment period and therefore could not be used.

Nevertheless, the reliance on the STS mortality risk calculator for MV Replacement results in several significant problems which affect the ability to adequately assess and interpret the meaning of data accumulated in a single arm registry trial. This is especially true when the output (STS score) is used as both the primary entry criteria for stratification of patients into "high risk" cohorts, and as the primary safety comparator to the observed mortality for the treatment received in the treatment arm (MitraClip). This quandary results from the fact that upon entering the exact same pre-operative variables for patient co-morbidities into the calculator, a markedly different STS score will result depending on the procedure chosen – in this case MV Repair or MV Replacement. In general, entering the same risk factors, the calculated STS score for MV replacement will be 50-100% higher than the STS score calculated for MV Repair. This has been confirmed by a review of the STS database looking at trends in MV surgery (Gammie JS et al. Ann Thorac Surg 2009;87:1431–9) where in the multivariable logistic regression model (incorporating 26 preoperative risk factors), risk-adjusted mortality was lower for mitral valve repair compared with replacement (odds ratio 0.52, 95% CI: 0.45 to 0.59, $p < 0.0001$; Table 3). Estimates were similar when indicator variables for hospital identity were included in the logistic model (odds ratio 0.54, 95% CI: 0.47 to 0.63, $p < 0.0001$). Results were also similar when risk adjustment was accomplished by propensity stratification instead of logistic regression: standardized mortality rate was 2.96% (95% CI: 2.72% to 3.20%) for MV replacement compared with 1.65% (95% CI: 1.49% to 1.80%) for MV repair ($p < 0.001$).

Table 9: Summary of the Logistic Model and Propensity Analysis for Mitral Valve repair and Replacement

Procedure	Number	Number of Deaths	Observed Mortality (%)	Odds Ratio (95% CI) MV Repair Versus MV Replacement		Risk Standardized Mortality, % (95% CI)[b]
				Unadjusted	Adjusted[a]	
MV repair	30,811	414	1.34	0.34 (0.30–0.39)	0.52 (0.45–0.59)	1.65 (1.49–1.80)
MV replacement	17,007	651	3.83	(reference)	(reference)	2.96 (2.72–3.20)

[a] Adjusted for 25 risk factors plus year of surgery ($p < 0.0001$ repair compared with replacement). [b] Directly adjusted across 10 propensity groups ($p < 0.001$).

CI = confidence interval; MV = mitral valve.

Although separate risk calculators for MV Repair and Replacement were not available to the Sponsor throughout the entire enrollment of these high risk registries, the sponsor's reliance on the calculator for MV Replacement is nevertheless problematic when attempting to use the output (STS score) for designating a "high risk" population of patients where the protocol defined patient population is elective and hemodynamically stable and the protocol defined valve pathology and morphology make them excellent candidates for MV Repair. Although a relative increase in patient mortality risk from that studied in the EVEREST II RCT can be assumed (patients are in a higher risk group), the target population necessary for determining benefit is not defined and, using the appropriate calculator, the mortality risk used to define the study populations entered into the registries may not adequately reflect the target population being sought for this device, i.e., "…*patients who have been determined by a cardiac surgeon to be too high risk for open mitral valve surgery."* Similar problems are introduced by the attempt to use the calculated mortality risk for MV Replacement as a primary outcome comparator for device treated patients who, as a result of strict inclusion and exclusion criteria, were excellent candidates for MV Repair.

The choice of valve repair vs. replacement in the hands of an experienced MV surgeon is dependent almost solely upon the valve pathology/morphology seen on pre-operative echocardiogram and assessed by direct inspection in the operating room. Although Gillinov et al. (J Thorac Cardiovasc Surg 2008;135:885-893) have shown that patients with irreparable mitral valves due to advanced degenerative morphology tend to be older with more co-morbidities, a causal link between age and/or co-morbidities and the inability to repair a mitral valve has not been suggested. They do note that *"when valvar pathology is so severe that repair is infeasible, valve replacement does not diminish long-term survival."* Given the strict valve criteria set forth in all of the EVEREST II studies to assure "clipability", the repair rate for these elective and hemodynamically stable patients, regardless of co-morbidities, is expected to approach 95% at the most experienced centers, and for the RCT arm of this trial, exceeded 87% (Guilinov AM et al. *J Thorac Cardiovasc Surg* 2008;135:885-893, Castillo JG et al. *J Thorac Cardiovasc Surg* 2012;144:308-12).

The effects of age on the ability to achieve a durable MV repair were examined utilizing linked STS and CMS databases. Badhwar et al. (Ann Thorac Surg 2012;94:1870-9), demonstrated that MV repair is a safe and durable long-term option for older patients. Survival restored to that of the normal age- and sex-matched general population suggests

repair may suppress the longitudinal impact of mitral regurgitation in the elderly. With regards to functional MR, the Cleveland Clinic group (Gillinov AM et al. J Thorac Cardiovasc Surg 2001;122:1125-1141) also noted that for Functional MR patients:

> *"Across all propensity-matched quintiles, mitral valve replacement resulted in a fairly uniform outcome, with poor long-term survival independent of patient status. In contrast, mitral valve repair produced a modulated result across the quintiles. Thus patients who underwent elective surgery and had a more stable condition derived a survival benefit from mitral valve repair."*

All available data for both functional and degenerative patients operated on electively suggest that whenever anatomically/morphologically feasible, MV Repair should be performed.

STS Score is calculated pre-operatively based upon a proprietary algorithm which calculates the average risk for the average patient and the average surgeon following the input of multiple clinically relevant patient risk factors (co-morbidities). Due to their rarity, these validated risk algorithms do not take into account certain anatomic risk factors such as porcelain aorta, previous mediastinitis or radiation, or disease states such as hepatic cirrhosis, and the assignment of mortality risk in patients with these factors must be supplemented with subjective judgment. The risk that an individual surgeon subjectively assigns to each of these conditions can vary widely based upon his/her experience and the expertise of the care team in the setting in which the surgeon operates.

The Inclusion/Exclusion related to mitral valve anatomic/pathologic/morphologic criteria for the two High Risk Registries (both HRR and Realism HR) are identical to those used for the EVEREST RCT and include:

INCLUSION Criteria

The primary regurgitant jet originates from malcoaptation of the A2 and P2 scallops of the MV. If a secondary jet exists, it must be considered clinically insignificant.

EXCLUSION Criteria

1. *MV orifice area < 4.0 cm2*
2. *If leaflet flail is present*
    a. *Width of the flail segment ≥15 mm, or*
    b. *Flail gap ≥10 mm*
3. *If leaflet tethering is present*
    a. *Coaptation depth > 11 mm*, or*
    b. *Vertical coaptation length is b2 mm*
4. *Severe mitral annular calcification*
5. *Leaflet anatomy that may preclude clip implantation, proper clip positioning on the leaflets, or sufficient reduction in MR. This may include the following:*
    i. *Evidence of calcification in the grasping area of the A2 and/or P2 scallops*
    ii. *Presence of a significant cleft of A2 or P2 scallops*
    iii. *More than 1 anatomic criteria dimensionally near the exclusion limits*
    iv. *Bileaflet flail or severe bileaflet prolapse*
    v. *Lack of both primary and secondary chordal support*

**criteria not repeated in HRR or Realism HR. All remaining Inclusion and Exclusion criteria regarding valvular anatomy/pathology are identical*

These strict valvular anatomic/pathologic/morphologic criteria in the design of the EVEREST RCT led to the following estimates of valve repairability by the investigators:

> *"It is estimated that 80% of patients undergoing surgery in the trial will undergo MV repair and the rest will undergo MV replacement. It is estimated that the incidence of death at 12 months is approximately 3% based on reported in-hospital mortality of 1.5% and 6% in patients undergoing surgical repair and replacement, respectively."*
>
> Mauri, L et al. The EVEREST II Trial: Design and rationale for a randomized study of the evalve mitraclip system compared with mitral valve surgery for mitral regurgitation. Am Heart J 2010;160:23-9

These predictions proved conservatively accurate. In the RCT surgical control arm, the following was observed:

- The ability to perform MV Repair was predicted pre-operatively in 75/80 surgical controls (93.75%);
- MV Repair was achieved in 69 of the 75 in whom it was predicted (92%); and
- For all patients randomized to surgery, MV Repair was performed in 69 of 80 (86.25%).

Hence, the relevant criteria used to determine the appropriate surgical procedure was consistent with repair being performed in over 85% of all patients in both the RCT and the two High Risk registries.

FDA concludes that the entirety of the inclusion and exclusion criteria did not uniformly specify a clinically "high risk" group. For instance, at least 5 patients (Table 10, below) were declared at "high risk" for surgery due to the presence of factors not captured by STS criteria (i.e., porcelain aorta, mediastinitis, radiation or cirrhosis), even though his/her overall STS score for Mitral Valve Repair by objective criteria may have been low (2-5%). These patients would, however, be assigned a minimal predicted surgical risk of 12%, with a potentially unlimited upper bound depending on subjective surgeon assessment of the added risk due to these conditions (i.e., surgeon may estimate a 40% mortality risk). Although, as noted above, these non-STS risk factors do affect the subjective determination of surgical risk, the assigned or estimated surgical risk, when used as a comparator, can significantly overstate the risk of a catheter based procedure in a patient undergoing a transseptal procedure using venous access. Furthermore, use of this individual potentially inflated value (and the similarly inflated values of other similarly situated patients) as individual inputs for the mean calculation of overall estimated predicted risk of mortality for the entire HRR cohort, serves to inflate the overall mean predicted risk of mortality used as the comparator for HRR (and REALISM HR) when compared to the more objectively defined HRR STS score.

The inclusion criteria also include the possibility of entering patients as "high risk" where such a classification is unjustified (actual STS score calculation <12% for MV Replacement, Table 10) based on all of the remaining objective STS criteria (inclusion criteria 4,5,6,7,9i-iv above). For example, a 76 year old patient with severe MR and NYHA Class III or Class IV heart failure symptoms and an ejection fraction (EF) of 35% could have been placed in this

cohort and assigned the minimal predicted mortality risk of 12% despite an STS operative risk for mitral valve repair of 1.6-2.2% (MV Replacement risk 2.6-3.7%). This results in risk factors (e.g., age, EF, NYHA Class, creatinine, previous operations) already included in the STS score being double- or triple-counted when used as part of the inclusion criteria for the HRR. In fact, for MV Repair, entering values into the STS risk calculator which are all worse than the thresholds cited for assignment of high risk by the protocol (Age 80, Cr 2.6, Class III NYHA symptoms, EF 30, previous CABG, first re-op,) results in an STS score for MV Repair of 6.4% (10.0% for MV Replacement; STS dataset 2.73) even though, for the purposes of estimating the "predicted surgical risk" used as the comparator, a minimal value of 12% would have been used. At least 25 patients of the 78 total HRR cohort (32%) were enrolled based on fulfillment of these criteria.

Table 10: Reasons for Including Patients Who Qualified Based on Pre-specified High Risk Co-Morbidities

Inclusion Criterion[a]	High Risk Registry (N = 30)	Concurrent Control (N = 12)
Porcelain aorta or mobile ascending aortic atheroma	6.7% (2/30)	0.0% (0/12)
Post-radiation mediastinum	3.3% (1/30)	8.3% (1/12)
Previous mediastinitis	0.0% (0/30)	0.0% (0/12)
Functional MR with LVEF < 40%	33.3% (10/30)	50.0% (6/12)
Over 75 years old with LVEF < 40%	20.0% (6/30)	8.3% (1/12)
Re-operation with patent grafts	53.5% (16/30)	50.0% (6/12)
Two or more prior chest surgeries	23.3% (7/30)	25.0% (3/12)
Hepatic cirrhosis	6.7% (2/30)	0.0% (0/12)
Three or more STS high risk factors[b]	10.0% (3/30)	8.3% (1/12)

[a] *Patients may be included in more than one category*
[b] *Creatinine > 2.5 mg/dL, prior chest surgery, age over 75, EF < 35%*

Alternatively, and similar to the situation where anatomic factors are used to arbitrarily estimate risk, the surgeon may have assigned an even higher value than 12%, based on a subjective estimate of the patient's surgical risk in his/her hands at his/her institution. Of the 78 HRR patients, 30 had a STS Score for MV repair that was likely substantially less than 12% but were judged "high risk" by a physician or by meeting one of the inclusion criteria. Of these, over half (16/30) were judged "high risk" at institutions that had relatively low volume mitral valve surgeons (<25 total MV operations per year). Other patients' status as high risk was determined without the benefit of a face-to-face surgical consult with the patient.

For the Primary Safety Endpoint, the sponsor compared the device's observed mortality rate to the "mean predicted surgical mortality rate" derived as follows:

- In patients with calculated STS score ≥ 12%, the actual STS Score is used, or
- In patients with calculated STS score < 12 who qualify due to the presence of one of the specific surgical risks listed in the inclusion criterion (reproduced above), the greater of the following is used:

- a minimal value of 12% for the surgical mortality risk estimate was specified; or
- the surgeon's subjective mortality risk estimate – no maximum limit for surgeon assigned risk estimate was provided.

FDA believes use of this highly subjective method for estimating surgical mortality rate to be used as a comparator to the observed device rate is not appropriate for the reasons discussed previously. Additional bias may have been introduced by the disparate levels of interventionalist/surgeon expertise inherent in the device-treated cohorts versus the pool of surgeons responsible for the results used for the comparator group. The device was placed by a highly select, experienced group of interventional cardiologists at select institutions. The STS Score includes over 90% of the surgeons in the US and represents the average results for the average surgeon at the average hospital. The STS database also includes all patients, while the device groups represent a highly selected subset of mitral valve patients (12% of those screened for the study). Furthermore, by its very nature as a database of patients who have actually undergone surgery, the surgical mortality risk of patients thought to be at too high a risk for surgery would not be part of the patient cohort used to develop or validate the STS calculation for predicted risk of mortality, making the STS score itself possibly unsuitable for use as an appropriate comparator.

Perhaps most importantly, examination of the STS datasheets provided by the sponsor in their PMA application showed that the mortality comparison to the STS database was made to the predicted risk of mortality for MV Replacement rather than MV Repair. Although this was the only calculation available to the sponsor at the time the majority of the patients were enrolled, this is an inappropriate comparator given the mitral valve anatomic inclusion/exclusion criteria which would predict a high (>90%) likelihood of repair in the overwhelming majority of patients in the hands of an appropriately experienced surgeon. As noted previously, for any set of STS risk factors, the predicted mortality for MV Replacement results in a calculated value of STS Score approximately 1.5 to 2.5 times higher than that calculated for MV Repair, resulting in a gross overstatement of the true surgical risk. Misuse of an inflated STS Score as a comparator cannot be justified, even if no misuse was intended. For instance, Pt #[redacted]/T-C (Appendix 25, page 43) has an STS score for replacement of 40.0% (dataset 2.52.1) where the identical risk factors for repair result in an STS score of 23.9% (dataset 2.73; replacement risk for same patient using this updated dataset is 34.9%). A patient with the average reported estimated surgical risk using only STS score for replacement (18.2%; Pt. #[redacted]/JVN, Appendix 25, page 381,dataset 2.52.1) has a STS score for repair of 7.43% using the currently available calculator (dataset 2.73).

STS risk calculator: http://sts.org/quality-research-patient-safety/quality/risk-calculator-and-models

FDA believes the term "high risk" merely reflects the relatively higher operative risk compared to the RCT group and reflects neither "too high risk for surgery' nor "inoperability" which are the designated target populations specified in the reconfigured Indications for Use statement submitted for this PMA. Judging operative risk and operability for these registry cohorts was highly subjective and substantially relied upon STS mortality risk calculation based on MV Replacement rather than the appropriate MV Repair. Judgment

regarding overall surgical risk assessment for this study may have been further complicated by the fact that only 50% of these patients were seen by cardiac surgeons, and that many of the surgeons participating in the registry study were not highly experienced mitral valve repair surgeons (<25 mitral valve repair procedures per year). The evaluation of surgical risk by non-surgeons or surgeons inexperienced in the routine evaluation of "high risk" mitral valve patients seriously hinders performance of accurate and reliable risk assessment, especially when subjective determinations of overall surgical risk are implemented. Importantly, the use of STS scores for MV Replacement instead of MV Repair to assess patient risk resulted in an overstatement of risk by a multiple of 1.5-2.5, and voids any ability to use either the STS Score or the estimated "predicted mortality risk" which is substantially derived from use of the flawed STS Score determination as a .means for assignment of "high risk" status or as a valid comparator for the ultimate determination of safety and effectiveness. Because of these reasons, it is difficult to accurately define the patients included in the HRR group as "high risk" and one cannot assume that these patients were inoperable. More importantly, the ability to determine whether or not the observed mortality for device therapy (Primary Safety Endpoint) was significantly lower than either the predicted surgical mortality (which relied substantially on use of the STS Score) or the objective STS Score is challenging..

Primary Safety Endpoint

For the Primary Safety Endpoint, the sponsor compared the device's observed 30-day mortality rate to the "mean predicted surgical mortality rate" which was derived as follows:

- In patients with calculated STS score ≥ 12%, the actual STS score is used, or
- In patients with calculated STS score < 12 who qualify due to the presence of one of the specific surgical risks listed in the inclusion criterion (reproduced above), the greater of the following is used:

    - a minimal value of 12% for the surgical mortality risk estimate was specified; or
    - the surgeon's subjective mortality risk estimate – no maximum limit for surgeon assigned risk estimate was provided.

The subjectively derived "predicted surgical risk" used by the sponsor for a determination of whether or not the Primary Safety Endpoint had been met for EVEREST had a mean of 18.2 +/- 8.0% (Table 11 below).

Table 11: EVEREST II HRR – Procedural Mortality at 30-days

	HRR (N = 78)
π_{Device}	7.7% (6/78)
95.472% UCB[a]	14.8%
$\pi_{Predicted}$	
Mean	18.2%
95% Two-sided Confidence Interval[b]	(16.4%, 20.0%)
Median	15.0%
95% Two-sided Confidence Interval[b]	(15.0%, 16.6%)

[a] UCB is based on the Clopper-Pearson method
[b] Confidence interval is calculated based on t-distribution and the confidence interval for the median is based on non-parametric methods.

Using the surgeon's subjective estimate of predicted surgical risk (i.e., 18.2%) for the comparator to the upper 95% confidence bound (14.8%) of the observed device mortality rate leads to the conclusion that the primary safety endpoint for the HRR registry was met. However, the opposite conclusion (Primary Safety Endpoint NOT met) would be reached using mean of the objectively derived STS score of 14.2% ± 8.2% for MV Replacement as the comparator. For the reasons discussed above, using the more appropriate comparator for MV Repair would also have resulted in failure to meet the endpoint as well, and would likely result in a mean STS score very close to the observed mortality rate for Device use (see example for patient [redacted]/JVN, above).

The 30-day rates for all MAEs including transfusion are shown in Table 12 below. Excluding transfusion, 12.8% of EVEREST HRR patients (10/78) had MAEs.

Table 12: Major Adverse Events at 30 Days and 12 Months High Risk Registry (N = 78)

Description of Event	Through 30 Days		Through 12 Months	
	% Patients (n/N)	# Events	% Patients (n/N)	# Events
Death[a]	7.7% (6/78)	6	23.1% (18/78)	18
Myocardial infarction	2.6% (2/78)	2	5.1% (4/78)	5
Re-operation for failed surgical repair or replacement	0.0% (0/78)	0	0.0% (0/78)	0
Non-elective cardiovascular surgery for adverse events	0.0% (0/78)	0	0.0% (0/78)	0
Stroke	2.6% (2/78)	2	2.6% (2/78)	2
Renal failure	3.8% (3/78)	3	6.4% (5/78)	5
Deep wound infection	0.0% (0/78)	0	0.0% (0/78)	0
Ventilation > 48 hours	2.6% (2/78)	2	2.6% (2/78)	2
GI complication requiring surgery	1.3% (1/78)	1	3.8% (3/78)	3
New onset of permanent AF	0.0% (0/78)	0	0.0% (0/78)	0
Septicemia	0.0% (0/78)	0	3.8% (3/78)	3
Transfusion of ≥ 2 units of blood	17.9% (14/78)	22	24.4% (19/78)	31
Total[b]	**26.9% (21/78)**	**38**	**42.3% (33/78)**	**69**

[a] *One additional patient [redacted] died between 30 days and 12 months and is reported in the analysis of the major effectiveness endpoint; however, this death is not included in this table since it occurred after the patient withdrew from the study.*

[b] *Some patients experienced more than one event, hence the total number of patients is not equal to the sum of the number of patients in each category.*

Primary Effectiveness Endpoint

FDA repeatedly informed the Sponsor that the HRR would not stand alone for approval, but rather could serve as adjunctive data used in support of approval based on the EVEREST II RCT data. The qualitative issues with these registry data are similar to the issues noted for the safety and effectiveness results of the RCT. The same concerns described for determination of safety apply. In addition, the Sponsor's decision to consider a reduction of MR to 2+ or less OR a reduction of MR by one grade as criteria for success for the effectiveness endpoint is regarded as inappropriate. Since the HRR is a single arm registry, there is no control group to serve as a comparator for effectiveness. While the Sponsor made an effort to address this limitation, the FDA notified the Sponsor of our concerns with their various proposals for a comparator for the HRR patients, including the following:

1. The FDA repeatedly informed the Sponsor that the proposed analysis, which compared safety to surgery and effectiveness to medical treatment, was inappropriate due to the

increased bias that would be introduced in favor of the treatment group in such an analysis.

2. The Sponsor identified a *post-hoc* "Concurrent Control" group (N=36) consisting of patients not meeting the inclusion/exclusion criteria for the HRR. The FDA did not agree that this was an appropriate control group since the patients, by definition, were not comparable given that they were excluded from enrollment in the HRR due to one or multiple reasons.

3. The sponsor explored comparison of the results of the HRR to literature. The sponsor stated that "*Although multiple publications report mortality during the peri-operative period following mitral valve surgery, the reported populations were generally not as high risk, or had significantly different co-morbidities compared to the patients considered for the EVEREST II HRR Study. Importantly, MR severity grade was not assigned by a core laboratory for any of the published reports" (PMA 100009 Amendment 17, Appendix 5 "Literature Comparators").*

It should be noted that the 30-day procedural mortality for the HRR cohort was 7.7% (6/78). One year mortality was 23% (18/78). Deaths adjudicated by the CEC showed that, overall, 61% (11/18) were either possibly or probably related to the device or procedure, including all 6 deaths within 30 days of the procedure. The limitations of defining the safety endpoint to be within 30 days is evident when one considers that 5 of the 18 deaths occurring between 30 days and 1 year were adjudicated by the CEC as being possibly or probably related to the device or procedure.

FDA Comment: The EVEREST II HRR was a heterogeneous, unblinded, single arm registry which collected an adjunct data set and is difficult to interpret in the setting of a failed pivotal randomized trial. FDA believes that concerns related to the High Risk Registry design and conduct are very significant. Using the HRR data alone, FDA is unable to determine that reasonable assurance of safety and effectiveness exists for the MitraClip CDS when used for the proposed indication for use in the designated target population.

7.2.3 REALISM – CONTINUED ACCESS PROTOCOL (CAP)

The High Risk arm of the REALISM study enrolled patients who met the same eligibility criteria as the EVEREST II HRR. The objective of the High Risk arm of the REALISM study was to collect registry data on the use of the MitraClip device in "high risk" patients. Due to the sequential and discontinuous nature of these studies, however, some differences in the respective populations can be identified. Unique to the CAP was the request for Compassionate Use in a significant number of patients where the basis of the request was specifically for inoperability. Although these patients were properly excluded from the overall analysis of high risk patients, it should be noted that this may have had the unintended consequence of effectively skewing the final composition of the REALISM HR population towards a less high risk population than was enrolled in the EVEREST HRR.

As with EVEREST HRR, REALISM HR patients were considered high surgical risk if their objectively calculated STS Score for Mitral Valve Replacement was ≥ 12%, or if the patient met one or more of the pre-specified surgical risk factors described in Table 13. Examination of this data provides further evidence that REALISM HR may have enrolled a less high risk cohort than the EVEREST HRR. Specifically, the proportion of patients enrolled in REALISM HR with an objectively determined STS Score of ≥ 12% was 37.7% (103/273) versus 61.5% (48/78) for EVEREST HRR. The remaining 170 patients were included due to the presence of one or more pre-specified surgical risk factors. Excluding patients in both cohorts who were declared high risk on the basis of one or more anatomic related factors (i.e., porcelain aorta, previous radiation or mediastinitis, hepatic cirrhosis; n=17 REALISM HR and n=5 HRR), 56% (153/273) of patients in REALISM HR were enrolled based on what was effectively double counting of risk factors already accounted for in the STS risk calculator, while these same criteria were used to enroll only 32% (25/78) of the EVEREST HRR cohort.

Table 13: Patients with Pre-specified Surgical Risk Factor (Patients with STS Mortality Risk <12%) REALISM High Risk and HRR Patients

Pre-Specified Surgical Risk Factor[a]	REALISM	HRR
Functional MR with LVEF < 40%	55.3% (94/170)	33.3% (10/30)
Re-operation with patent grafts	48.2% (82/170)	53.5% (16/30)
Two or more prior chest surgeries	20.0% (34/170)	23.3% (7/30)
Over 75 years old with LVEF < 40%	26.5% (45/170)	20.0% (6/30)
Three or more STS high risk factors[b]	7.6% (13/170)	10.0% (3/30)
Hepatic cirrhosis	2.4% (4/170)	6.7% (2/30)
Porcelain aorta or mobile ascending aortic atheroma	2.9% (5/170)	6.7% (2/30)
Post-radiation mediastinum	4.1% (7/170)	3.3% (1/30)
Previous mediastinitis	0.6% (1/170)	0.0% (0/30)

[a] *Patients may be included in more than one category*
[b] *Creatinine > 2.5 mg/dL, prior chest surgery, age over 75, EF < 35%*

These proportional disparities in the composition of the REALISM HR and EVEREST HRR cohorts resulted in a large and significant difference in the STS score for the REALISM HR and EVEREST HRR cohorts (10.5% vs, 14.2%, respectively). It is consequently difficult to interpret the full clinical implications of the observed procedural mortality in Table 14 below.

Table 14: Procedural Mortality REALISM HR Patients (N = 273)

π_{Device}	4.0% (11/273)
95% two-sided CI[a]	(2.0%, 7.1%)
$\pi_{Predicted}$	
Mean	10.5%
95% two-sided CI[b]	(9.6%, 11.4%)
Median	8.9%
95% two-sided CI[b]	(7.9%, 10.3%)

[a] *CI: Confidence Interval is based on the Clopper-Pearson method*
[b] *Confidence interval is calculated based on a t-distribution and the confidence interval for the median is based on non-parametric methods*

Finally, as was noted with EVEREST HRR, any true assessment of predicted surgical risk in the REALISM HR cohort is further compounded by the improper use of the STS Score for Replacement rather than Repair.

The overall incidence of procedural MAEs for the REALISM HR cohort is summarized below.

Table 15: Major Adverse Events at 30 Days and 12 Months REALISM HR Patients (N = 273)

Description of Event	**MAE at 30 Days**		**MAE at 12 Months**	
	% Patients (n/N)	**# Events**	**% Patients (n/N)**	**# Events**
Death	4.0% (11/273)	11	22.7% (62/273)	62
Myocardial infarction	0.7% (2/273)	2	1.5% (4/273)	5
Re-operation for failed surgical repair or replacement	0.0% (0/273)	0	0.0% (0/273)	0
Non-elective cardiovascular surgery for adverse events	0.4% (1/273)	1	0.4% (1/273)	1
Stroke	2.6% (7/273)	7	3.7% (10/273)	10
Renal Failure	1.1% (3/273)	3	5.1% (14/273)	14
Deep Wound Infection	0.0% (0/273)	0	0.0% (0/273)	0
Ventilation > 48 hours	2.9% (8/273)	8	6.2% (17/273)	20
GI complication requiring surgery	0.0% (0/273)	0	0.7% (2/273)	2
New onset of permanent AF	0.4% (1/273)	1	0.4% (1/273)	1
Septicemia	1.1% (3/273)	3	4.4% (12/273)	13
Transfusion of ≥ 2 units of blood	12.1% (33/273)	35	22.0% (60/273)	79
Total[a]	**16.5% (45/273)**	**71**	**36.3% (99/273)**	**207**
Total[a] (Excluding Transfusions ≥ 2 units)	**8.1% (22/273)**	**36**	**27.1% (74/273)**	**128**

[a] *Total number of patients may not equal the sum of patients in each row since one patient may experience multiple events.*

It should be noted that the number of deaths at 12 months in the above table includes 3 subjects that died after 365 days.

The lower device mortality in the REALISM HR registry (4.0%) versus the EVEREST HRR cohort (7.7%) along with the overall lower rates of procedural (30-day) MAEs (8.1% in REALISM HR vs. 12.8% in EVEREST HRR, excluding transfusions) despite use of identical inclusion/exclusion criteria further supports the above noted observations about the relative differences in overall cohort surgical risk composition, and also suggests the presence of at least two additional factors that might favorably influence results in sequential discontinuous trials:

- Selection bias; and
- Improved operator experience and judgment.

FDA Comment: REALISM HR is a continued access protocol study that was not intended to be used as a pivotal dataset and is difficult to interpret. FDA believes that the number and severity of concerns related to the REALISM High Risk registry design and conduct are identical in nature and severity to those identified for the EVEREST HRR. These limitations make evaluation of whether there is a reasonable assurance of safety and effectiveness for use of the MitraClip device for the proposed indication and designated population challenging.

8 OUS EXPERIENCE: ACCESS-EU

The MitraClip was approved in Europe (received CE Mark) in March 2008. The ACCESS – EU study was a 2-phased prospective, observational, multicenter study of the use of MitraClip with a duration of 12 months. The Sponsor has provided the information that is summarized here.

ACCESS-EU had 2 Phases. Phase I was composed of patients who received the MitraClip and two comparator groups: One control group is a medically managed group and the other is a group who received mitral valve surgery. Phase II included patients who received the MitraClip as part of clinical practice. It is uncertain whether the Sponsor is still entering patients into Phase II. Not all patients implanted have been enrolled in the registry. The Sponsor states that the 567 is a subset of the all Europe commercial implants. No reasons are provided that inform how the 567 were selected to enter the registry. The "comparator group" of medically managed patients is not provided in this report.

The patient profile was >70 years old and primarily ischemic with functional MR. The comorbidities observed are the typical ones seen in ischemic patients as well. The NYHA class was advanced with mostly moderate to severe MR, but the patients in whom hemodynamics are reported are only mildly-moderately abnormal. It should be noted that more than 50% of the patients have missing data on functional assessment, NYHA, and QOL, and nearly 240/567 measurements of LV diameters were not reported. It is also unclear how the level of MR described is associated with a small ventricle at baseline (LVEDD of 4.6 cm. After MitraClip implantation, there is a reduction in MR severity observed.

However, no data are provided for the LVEDD post-procedure. Therefore, for the data reported, there exist inconsistencies between the severity of NYHA, MR and the hemodynamics.

In Phase I patients, the 30-day mortality was 3.4%. The mean logistic EuroScore for predicted surgical risk at baseline was 23.1% in these patients. Other major adverse events also occurred at low rates (<2%). The Kaplan-Meier freedom from mortality rate at 12 months is reported to be 81.7%. The Sponsor emphasizes that a large percentage of patients had high EuroScores, a measure of surgical risk used in Europe and developed primarily for CABG surgery. The Sponsor further states that the patients that died had the higher EuroScores and therefore were "high risk." The EuroScore, a measure of risk of surgery is primarily used in Europe and was developed primarily for CABG surgery. No information was provided as to whether any of the patients were seen by a surgeon and deemed high risk for surgery.

The reasons for the large amount of missing data are not provided.

FDA Comment: FDA believes it is difficult to draw conclusions from a highly select patient group in which the selection process for entry into the dataset is unknown. Lack of background therapy or a comparison group which undergoes surgery (an objective of Phase I), make the interpretation of these data difficult. Therefore, this ACCESS-EU group provides little added information on the patients defined by the sponsor's proposed indication for use.

9 *POST-HOC* ADDITIONAL DATA ANALYSES

9.1 INTEGRATED HIGH SURGICAL RISK COHORT

9.1.1 Background

The individual shortcomings with regards to the design and conduct of the separately and sequentially enrolled exploratory HRR and REALISM High Risk Registries were discussed previously in this document. Each has undergone extensive *post-hoc* analysis yielding results that are compelling for hypothesis generation and supportive for the performance of an appropriately-designed, well-conducted randomized controlled pivotal trial. Considered separately, however, FDA is concerned that neither of these two data sets offers the valid scientific evidence necessary to allow a determination of reasonable assurance of safety and effectiveness that is required for approval of this first-of-a-kind device designed to treat severe MR in high risk patients.

Recognizing that uncontrolled outcome data from individual registries required further support to form the basis of an argument for approval based on a *post-hoc* reconfiguration of both the Indication for Use and target population, the Sponsor pooled these two sequentially enrolled registry cohorts to form the "Integrated High Surgical Risk Cohort." The Integrated High Surgical Risk Cohort is comprised of 351 patients, including:

- 78 patients who were enrolled in the EVEREST II HRR and have completed 3 years of follow-up, and
- 273 patients who were subsequently enrolled in the High risk arm of the REALISM continued access study (273 with one year follow-up – includes 133 patient with 2-year follow-up plus 140 additional patients with 1-year follow-up).

Table 16: Cohorts of patients pooled to form the Full Integrated High Surgical Risk cohort

Table 1 reference	Study/Cohort	# of Pts	30-day	12-month	24-month	3-year
III	EVEREST II High Risk Registry	78	N = 72 D = 6 W = 0 M = 0	N = 56 D = 18 W = 3 M = 1	N = 44 D = 26 W = 7 M = 1	N = 39 D = 31 W = 7 M = 1
IV	EVEREST II REALISM High Risk	133 (273)	N = 124 D = 6 W = 2 M = 1	N = 93 D = 32 W = 5 M = 3	N = 78 D = 47 W = 7 M = 1	
	Additional EVEREST II REALISM High Risk	140	N = 128 D = 6 W = 3 M = 3	N = 98 D = 30 W = 8 M = 4		

The goal of combining these registry cohorts into the single Integrated High Surgical Risk Cohort was to support a newly hypothesized Indication for Use that was configured based on the known combined clinical outcomes of these pooled registry cohorts. This revised *post-hoc* Indication for Use statement presented for consideration by the Sponsor also includes a new target population and is summarized below:

The MitraClip Clip Delivery System is indicated for the percutaneous reduction of significant symptomatic mitral regurgitation (MR≥3+) in patients who have been determined by a cardiac surgeon to be too high risk for open mitral valve surgery and in whom existing co-morbidities would not preclude the expected benefit from correction of the mitral regurgitation.

FDA agrees this is a compelling population. By definition, no current effective interventional or surgical therapy exists for this unfortunate and symptomatic group of patients. However, the effectiveness of invasive therapies in the treatment of these patients has not been demonstrated. Great enthusiasm for the role of surgical therapy followed initial reports by Bolling, et al.,[1,2] demonstrating the ability to reliably eliminate or reduce MR with low operative mortality (5%) and initial symptomatic improvement in patients with functional MR. However, subsequent follow-up study of these same patients by Wu, et al.,[3] showed no benefit in 5-year mortality when compared to propensity matched patients treated at the same institution over the same period of time (1995-2002). Although the patients treated in that study had demonstrably worse LV function (mean EF 20-23%, mean LVEDD 62-65mm) than those currently being considered, a low surgical mortality risk (5%), similar to those seen in the current era with MitraClip device treatment in the high risk populations (7.7% EVEREST HRR, 4.0% REALISM HR) was achieved. The key finding was that the near complete elimination of MR with surgery did not result in a survival benefit at 5 years when compared to medical therapy. Although the desire to offer safe and effective therapy to these patients persists unabated, the proper role of medical therapy and/or additional interventional/surgical procedures remains uncertain, and the degree of MR relief required to effect clinically meaningful and durable improvement in symptomatic patients with no other treatment options is unknown.

1. Bolling SF et al. Intermediate-term outcome of mitral reconstruction in cardiomyopathy. J Thorac Cardiovasc Surg 1998;115:381– 6.
2. Bach DS, Bolling SF. Improvement following correction of secondarymitral regurgitation in end-stage cardiomyopathy with mitral annuloplasty.Am J Cardiol 1996;78:966 –9.
3. Wu, AH et al. Impact of Mitral Valve Annuloplasty on Mortality Risk in Patients With Mitral Regurgitation and Left Ventricular Systolic Dysfunction. J Am Coll Cardiol 2005;45:381–7

Since neither registry contained a suitable comparator that would provide the definitive valid scientific evidence necessary to allow reasonable assurance of safety and effectiveness for the specified indication and target population noted above, additional *post-hoc* analyses were conducted. The sponsor compared the observed outcomes of this pooled Integrated High Surgical Risk Cohort to the results of medically treated patients extracted from the Duke Database (Duke Cohort), via propensity score matching methodology, and from the Ohio State University Cardiac Database.

The Sponsor proposes that these *post-hoc* analyses, rather than being regarded as hypothesis generating, are sufficient to provide the valid scientific evidence necessary to show reasonable assurance of safety and effectiveness for their revised Indication for Use and target population. In order to support this position, the Sponsor must be able to demonstrate that the following two separate data manipulations are both appropriate and successful:

1. Pooling of the HRR and REALISM HR cohorts to form the Integrated High Surgical Risk Cohort; and
2. Successful propensity score matching required for comparison of outcomes between the Integrated High Surgical Risk Cohort and Duke Cohort.

FDA has examined the appropriateness and success of each step required to make the Sponsor's proposed comparative analyses and has concluded that significant problems exist at each step.

9.1.2 Inclusion/Exclusion Criteria

Pooling of the HRR and REALISM HR populations require that reasonably similar patients are enrolled in each study. Although these were sequentially enrolled registries, the same Inclusion and Exclusion criteria were used for each as summarized below:

The **Inclusion Criteria** for both the EVEREST II HRR and REALISM HR patient cohorts were:

1. Predicted procedural mortality risk calculated using the STS surgical risk calculator of ≥12%, or if in the judgment of the surgeon Investigator, the patient is considered a high risk surgical candidate due to the presence of one of the following co-morbidities:

    a. Porcelain aorta or mobile ascending aortic atheroma
    b. Post-radiation mediastinum
    c. Previous mediastinitis
    d. Functional MR with EF < 40%
    e. Over 75 years old with EF < 40%

f. Prior re-operation with patent grafts
g. Two or more prior chest surgeries
h. Hepatic cirrhosis
i. Three or more of the following STS high risk factors:
    i. Creatinine > 2.5 mg/dL
    ii. Prior chest surgery
    iii. Age over 75
    iv. EF < 35%

2. Symptomatic moderate-to-severe (3+) or severe (4+) chronic mitral valve regurgitation, and in the judgment of the investigator, intervention to reduce MR is likely to provide symptomatic relief for the patient.

3. The primary regurgitant jet originates from malcoaptation of the A2 and P2 scallops of the mitral valve.

4. Transseptal catheterization is determined to be feasible by the treating physician.

5. Age 18 years or older.

6. Male or Female. Female subjects of childbearing potential must have a negative pregnancy test within seven (7) days before the procedure.

7. American Society of Anesthesiologists physical status classification of ASA IV or lower.

8. The subject or the subject's legal representative has been informed of the nature of the study and agrees to its provisions and has provided written informed consent as approved by the Institutional Review Board of the respective clinical site.

9. The subject and the treating physician agree that the subject will return for all required post-procedure follow-up visits

The **Exclusion criteria** for each study are presented below:

1. Severe left ventricular dysfunction, defined as an ejection fraction < 20%, and/or end-systolic dimension > 60 mm.

2. Mitral valve orifice area <4.0cm^2.

3. Mitral valve anatomical exclusions

    a. If leaflet flail is present (degenerative MR):

        i. Flail Width: flail segment width greater than or equal to 15 mm
        ii. Flail Gap: the flail gap is greater than or equal to 10 mm

    b. If leaflet tethering is present (functional MR):

        i. Coaptation Length: the vertical coaptation length is less than 2 mm

    c. Leaflet anatomy which may preclude device implantation, proper MitraClip device positioning on the leaflets or sufficient reduction in MR. This may include:

        i. Evidence of calcification in the grasping area of the A2 and/or P2 scallops
        ii. Presence of a significant cleft of A2 or P2 scallops
        iii. More than one anatomic criteria dimensionally *near* the exclusion limits

iv. Bileaflet flail or severe bileaflet prolapse
v. Lack of both primary and secondary chordal support

4. Evidence of an acute myocardial infarction in the prior 2 weeks before the intended treatment (defined as: Q wave or non-Q wave infarction having CK enzymes ≥2X the upper laboratory normal limit with the presence of a CK-MB elevated above the institution's upper limit of normal).

5. Hemodynamic instability defined as systolic pressure <90mmHg without afterload reduction or cardiogenic shock or the need for inotropic support or intra-aortic balloon pump.

6. Need for emergent or urgent surgery for any reason.

7. Prior mitral valve leaflet surgery or any currently implanted mechanical prosthetic mitral valve.

8. Echocardiographic evidence of intracardiac mass, thrombus, or vegetation.

9. Active endocarditis or active rheumatic heart disease or leaflets degenerated from rheumatic disease.

10. History of bleeding diathesis or coagulopathy or subject will refuse blood transfusions.

11. Active infections requiring current antibiotic therapy; may enroll 2 weeks post discontinuation of antibiotic therapy. Patients must be free from infection prior to treatment. Any required dental work should be completed a minimum of 3 weeks prior to treatment.

12. Intravenous drug abuse or suspected inability to adhere to follow-up.

13. A known hypersensitivity or contraindication to study or procedure medications which cannot be adequately managed medically.

14. Patients for whom transesophageal echocardiography (TEE) is contraindicated

15. In the judgment of the Investigator, the femoral vein cannot accommodate a 24F catheter or, if patient has ipsilateral deep vein thrombosis.

16. In the judgment of the Investigator, patients for whom the presence of a permanent pacemaker or pacing leads would interfere with placement of the test device or the placement of the test device would disrupt the leads.

17. Currently participating in an investigational drug or another device study that has not completed the primary endpoint or that clinically interferes with the current study endpoints. [Note: Trials requiring extended follow-up for products that were investigational, but have since become commercially available, are not considered investigational trials.]

FDA Comment: It must be re-stated that FDA believes the term "high risk" merely reflects the relatively higher operative risk compared to the RCT group and reflects neither "too high risk for surgery' nor "inoperability" which are the designated target populations specified in the reconfigured Indications for Use statement submitted for this PMA. The shortcomings regarding the inclusion/exclusion criteria, lack of appropriate surgeon involvement in assessment of surgical risk, the influence of subjectivity and "double counting" of STS risk factors for determining "high risk," as well as the inappropriate use of the STS score for MV Replacement rather than Repair, have been thoroughly discussed in the HRR section and apply equally to REALISM HR and the pooled Integrated High Risk Cohort.

9.1.3 Poolability Analysis

Enrollment and execution of the HRR and REALISM HR registries were sequential and discontinuous, that is, enrollment did not overlap and there was some time elapsed between the two registries' enrollment periods. The Sponsor considered the evaluation of poolability of patient data from these studies reasonable because no significant changes in device design or clinical eligibility criteria were made between the two studies.

Table 17 shows the Baseline Characteristics used to assess poolability of the HRR (78) and REALISM HR (273) cohorts

Table 17: REALISM High Risk (N=273) and EVEREST II HRR (N=78) Patients – Assessment of Poolability

Characteristic	REALISM (N = 273)	HRR (N = 78)	p-value
Age, years			
Mean ± SD (N)	75.5 ± 10.7 (273)	76.7 ± 9.8 (78)	0.353
Patients over 75 years of age, % (n/N)	57.1% (156/273)	61.5% (48/78)	0.517
Male Gender, % (n/N)	60.4% (165/273)	62.8% (49/78)	0.793
Body Mass Index (kg/m2)			
Mean ± SD (N)	26.9 ± 12.9 (273)	26.6 ± 5.0 (78)	0.749
Atrial Fibrillation, % (n/N)	70.5% (172/244)	61.6% (45/73)	0.155
Diabetes, % (n/N)	39.0% (106/272)	41.0% (32/78)	0.793
Myocardial Infarction, % (n/N)	49.3% (134/272)	55.8% (43/77)	0.366
COPD, % (n/N)			0.223
With home O_2	11.4% (31/272)	10.3% (8/78)	
Without home O_2	15.8% (43/272)	24.4% (19/78)	
None	72.8% (198/272)	65.4% (51/78)	
Stroke, % (n/N)	13.6% (37/273)	10.3% (8/78)	0.565
NYHA Class III/IV, % (n/N)	83.5% (228/273)	89.7% (70/78)	0.211
Previous Cardiovascular Surgery, % (n/N)	60.1% (164/273)	59.0% (46/78)	0.896
Functional MR Etiology, % (n/N)	73.3% (200/273)	59.0% (46/78)	0.018
LV Internal Dimensions, systole (cm)			
Mean ± SD (N)	4.5 ± 1.1 (245)	3.9 ± 1.1 (78)	< 0.0001
LV Ejection Fraction, % (n/N)			
Mean ± SD (N)	45.2 ± 13.6 (240)	54.4 ± 13.7 (78)	< 0.0001

[3] Sample sizes or denominators smaller than that in the header reflect missing data.

This analysis revealed significant differences between the REALISM HR and HRR cohorts in critically important objective etiologic and morphologic variables that were measured:

- MR etiology (unequal distribution of Functional vs. Degenerative MR between REALISM HR and HRR, $p < 0.02$);
- Baseline LVEF (45% REALISM vs. 54% HRR, $p < 0.0001$); and
- Baseline LVIDs (4.5cm REALISM vs. 3.9cm HRR, $p < 0.0001$).

As a result, the patients from the REALISM HR registry, which started enrollment after completion of HRR enrollment, represent a morphologically distinct subset of high risk patients with severe MR consisting of a significantly higher proportion of patients with functional vs. degenerative MR, as well as patients with overall worse LV function (lower EF) and more dilated ventricles (LVIDs). These differences in key morphologic variables

imply that there are significant clinical differences in the two populations of patients represented by REALISM HR and EVEREST II HRR.

It should also be noted that the pre-specified criteria summarized in the above table (Table 3), though significant, omit comparisons between several other important baseline characteristics which contribute to the STS predicted risk of mortality (PROM) including the incidences of heart failure, CAD, angina, cerebrovascular disease, and moderate to severe kidney disease, etc. The Sponsor does provide descriptive statistics for all baseline characteristics in Table 18, comparing REALISM HR (N = 273) and HRR (N = 78), but does not include p-values for these comparisons.

Table 18: REALISM High Risk and EVEREST II HRR – Key Demographics

Characteristic[a] % (n/N)	**REALISM High Risk (N = 273)**	**HRR (N = 78)**
Age, years		
Mean±SD (N)	75.5 ±10.7 (273)	76.7±9.8 (78)
Patients over 75 years of age	57.1% (156/273)	61.5% (48/78)
Male Gender	60.4% (165/273)	62.8% (49/78)
Body Mass Index, kg/m^2 Mean±SD (N)	26.9 ±12.9 (273)	26.6±5.0 (78)
Congestive Heart Failure	97.4% (266/273)	100.0% (78/78)
Coronary Artery Disease	81.7% (223/273)	84.2% (64/76)
Myocardial infarction	49.3% (134/272)	55.8% (43/77)
Angina	48.8% (124/254)	57.1% (44/77)
Atrial fibrillation	70.5% (172/244)	61.6% (45/73)
Cerebrovascular disease	21.2% (58/273)	17.9% (14/78)
Peripheral vascular disease	19.2% (52/271)	18.2% (14/77)
Cardiomyopathy	59.3% (162/273)	51.3% (40/78)
Hypertension	89.4% (244/273)	89.7% (70/78)
Diabetes	39.0% (106/272)	41.0% (32/78)
Moderate to Severe Renal Disease	32.6% (89/273)	23.1% (18/78)
Previous Cardiovascular Surgery, % (n/N)		
Coronary artery bypass graft	56.0% (153/273)	55.1% (43/78)
Aortic valve surgery	5.5% (15/273)	3.8% (3/78)
Tricuspid valve surgery	0.4% (1/273)	0.0% (0/78)
Other cardiac surgery	9.2% (25/273)	7.7% (6/78)
Previous PCI, % (n/N)	53.1% (145/273)	38.5% (30/78)
Cardiac Rhythm Device Implant, % (n/N)		
Pacemaker	14.4% (38/263)	22.1% (17/77)
ICD	29.7% (78/263)	13.0% (10/77)
NYHA Functional Class, % (n /N)		
I	3.3% (9/273)	0.0% (0/78)
II	13.2% (36/273)	10.3% (8/78)
III	61.9% (169/273)	60.3% (47/78)
IV	21.6% (59/273)	29.5% (23/78)
LV Ejection Fraction, % Mean±SD (N)	45.2 ±13.6 (240)	54.4±13.7 (78)
LV Internal Diameter, systole, cm Mean±SD (N)	4.5 ±1.1 (245)	3.9±1.1 (78)
STS Predicted Mortality Risk Score[b]	10.5% ± 7.4% (273)	14.2% ± 8.2% (78)

[a] Sample sizes or denominators smaller than 273 reflect missing data
[b] Predicted mortality was calculated using STS version 2.52 for EVEREST II HRR and STS version 2.61 for REALISM

In addition to the significant divergences previously noted in key baseline morphologic descriptors between the REALISM High Risk and HRR cohorts (Table 17, above), the most objective evidence available, the STS score, suggests important clinical differences may be present in these two populations (Table 18 above; REALISM High Risk 10.5% ±7.4% vs. HRR 14.2% ± 8.2%). The true magnitude/significance of the differences in STS score for these two cohorts remains unknown due to the decision to utilize the STS risk calculation for MV Replacement rather than MV Repair. It should be noted that FDA believes a more appropriate comparison would be to predicted rate of mortality for MV Repair.

Based on the previously noted baseline etiologic and morphologic differences which would intuitively seem to clinically favor HRR patients (lower incidence of functional MR etiology, higher baseline EF, and lower baseline LVIDs), the divergence in reported STS scores is counterintuitive, and occurred despite a higher percentage of patients in the REALISM HR cohort being designated as positive for factors having a substantial negative influence on STS score such as the higher incidence of moderate to severe kidney disease (REALISM HR 33% vs. HRR 23%), lower ejection fraction (REALISM HR 45% mean vs. HRR 54% mean,), higher pre-procedural AF (REALISM HR 71% vs. HRR 62%), and higher incidence of previous PCI (REALISM HR 53% vs. HRR 39%). In addition, cardiomyopathy and ICD use, which are not used in STS score calculations but may also suggest a less optimal baseline state of ventricular function and negatively affect surgical risk, were both higher in the REALISM HR cohort as well.

The pooled Integrated High Surgical Risk Cohort demographic characteristics are summarized below:

Table 19: Baseline Demographic Characteristics - Integrated High Surgical Risk Cohort

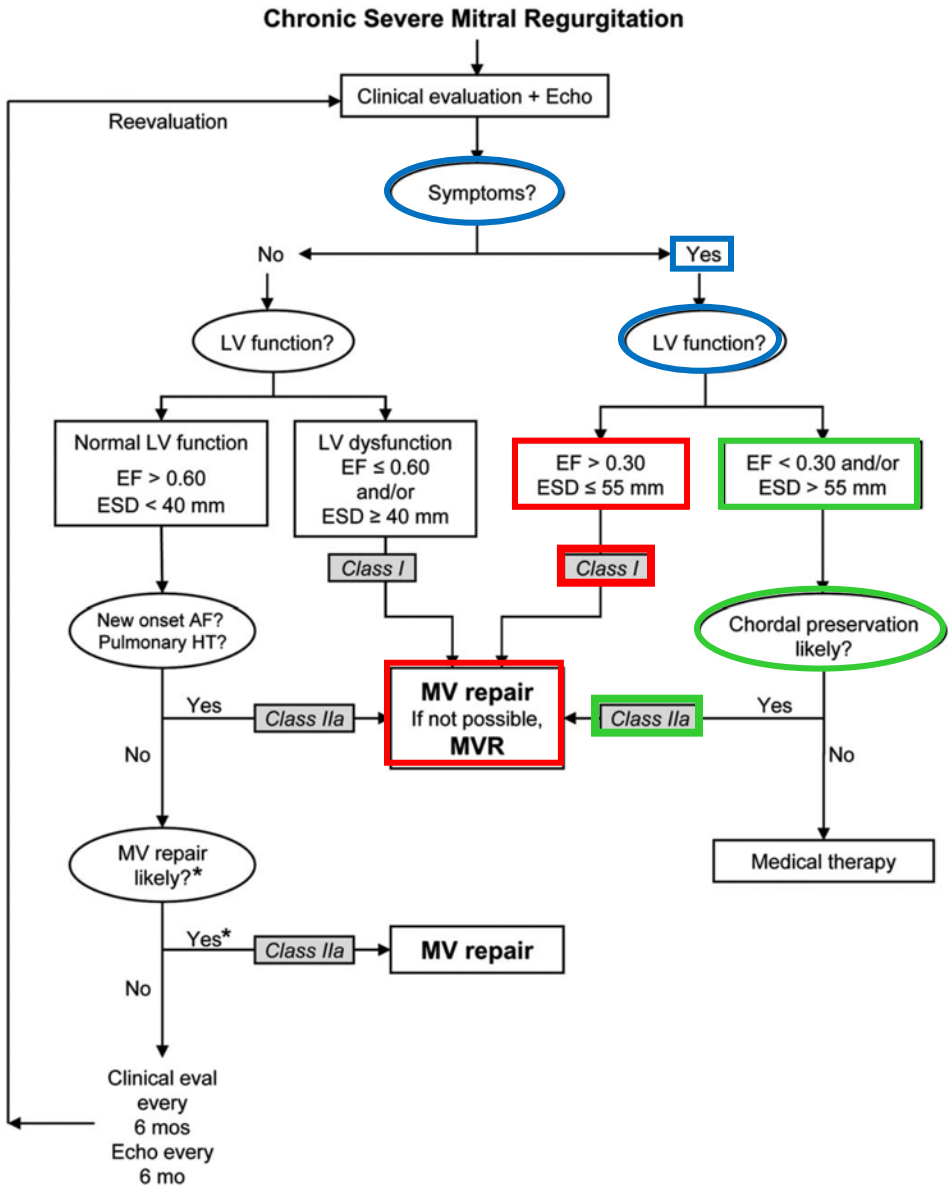
Characteristic[a] % (n/N)	**Integrated High Surgical Risk Cohort (N=351)**
Age, years	
Mean±SD (N)	75.7±10.5 (351)
Patients over 75 years of age	58.1% (204/351)
Male Gender	61.0% (214/351)
Body Mass Index, kg/m^2 Mean±SD (N)	26.9±11.6 (351)
Congestive Heart Failure	98.0% (344/351)
Coronary Artery Disease	82.2% (287/349)
Myocardial infarction	50.7% (177/349)
Angina	50.8% (168/331)
Atrial fibrillation	68.5% (217/317)
Cerebrovascular disease	20.5% (72/351)
Stroke	12.8% (45/351)
Peripheral vascular disease	19.0% (66/348)
Cardiomyopathy	57.5% (202/351)
COPD	28.9% (101/350)
Hypertension	89.5% (314/351)
Diabetes	39.4% (138/350)
Moderate to Severe Renal Disease	30.5% (107/351)

Characteristic[a] % (n/N)	Integrated High Surgical Risk Cohort (N=351)
Previous Cardiovascular Surgery, % (n/N)	59.8% (210/351)
Coronary artery bypass graft	55.8% (196/351)
Aortic valve surgery	5.1% (18/351)
Tricuspid valve surgery	0.3% (1/351)
Other cardiac surgery	9.1% (32/351)
Previous PCI, % (n/N)	49.9% (175/331)
Cardiac Rhythm Device Implant, % (n/N)	
Pacemaker	20.3% (69/340)
ICD	30.0% (102/340)
NYHA Functional Class, % (n /N)	
I	2.6% (9/351)
II	12.5% (44/351)
III	61.5% (216/351)
IV	23.4% (82/351)
LV Ejection Fraction, % Mean±SD (N)	47.5±14.2 (318)
LV Internal Diameter, systole, cm Mean±SD (N)	4.4±1.1 (323)

[a] Sample sizes or denominators smaller than 351 reflect missing data

In the absence of factors predictive of a prohibitive risk for operative mortality, symptomatic patients with severe MR and the LVEF and left ventricular end systolic dimensions (LVESD) exhibited by this pooled cohort would normally fall within the Class I indications for surgery according to ACC/AHA Guidelines summarized in Figure 6 below (blue to red path). Because the anatomic valvular anatomy defined by the inclusion/exclusion criteria for these two cohorts make chordal preservation almost certain, even symtomatic patients falling outside of these EF (< 30) and ESD (>55mm) parameters would receive a recommendation for surgical therapy (Class IIA – blue to green path).

Figure 6: ACC/AHA treatment Guidelines for patients wih Mitral Regurgitation



Bonow, RO, et al. ACC/AHA 2006 Guidelines for the Management of Patients With Valvular Heart Disease JACC 2006; 48: e1–148

For the Integrated High Surgical Risk Cohort, 200 patients were entered despite an objective calculated STS score for mitral valve replacement of less than 12. Anatomic criteria (procelain aorta, mediastinitis, radiation, and cirrhosis) accounted for 22 of these, with the remainder (n=178/351), over one half of all patients, entered due to the assignment of "high risk" based on the use of criteria already accounted for (both alone and in combination) in the STS risk score calculation. The critical shortcomings of this methodology are substantial and are fully discussed in the High Risk Resgistry portion of this document. It is important to note, however, that all of the non-anatomic variables noted below (LVEF, age, re-operation and number of prior surgeries, creatinine, ischemic CAD) are acounted for by the STS risk

calculator. Even a patient triggering all of them would have a STS score for MV repair of 6%, a mortality risk unlikely to be considered inoperable for surgeons/centers experienced in the care of these patients.

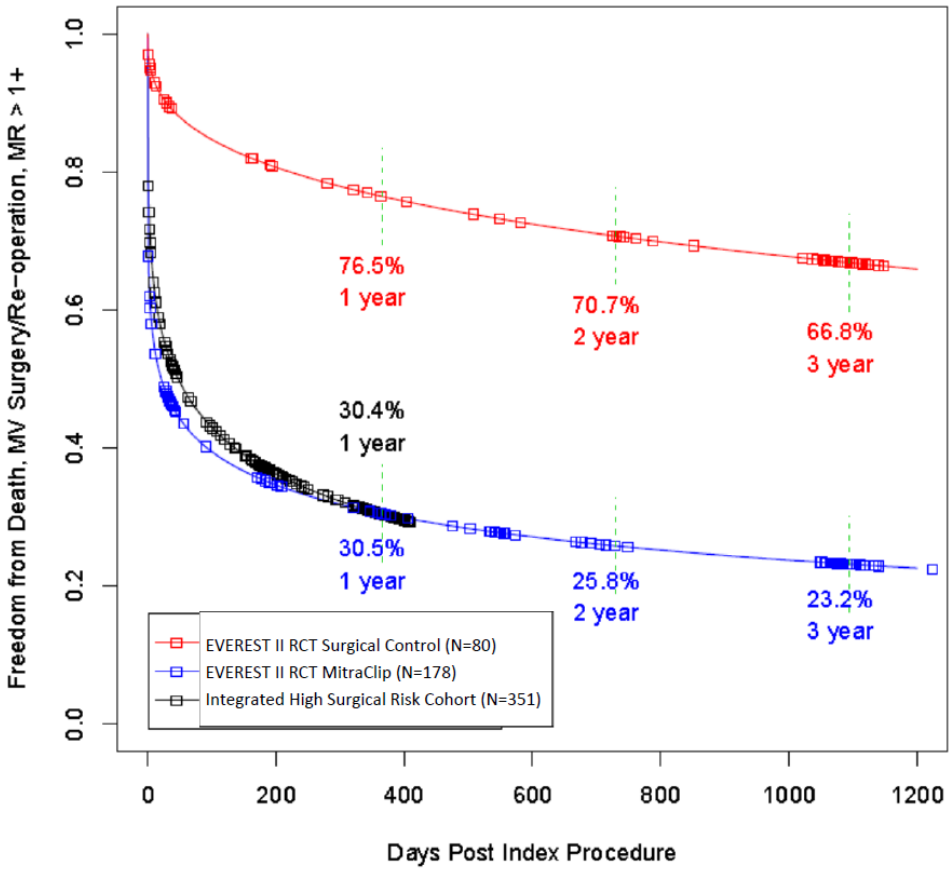
Table 20: EVEREST Integrated High Surgical Risk Cohort for Patients with Pre-specified Surgical Risk Factors (STS Mortality Risk <12%)

Pre-Specified Surgical Risk Factor[a]	EVEREST Integrated High Risk Cohort (N=200)
Functional MR with LVEF < 40%	52.0% (104/200)
Re-operation with patent grafts	49.0% (98/200)
Two or more prior chest surgeries	20.5% (41/200)
Over 75 years old with LVEF < 40%	25.5% (51/200)
Three or more STS high risk factors[b]	8.0% (16/200)
Hepatic cirrhosis	3.0% (6/200)
Porcelain aorta or mobile ascending aortic atheroma	3.5% (7/200)
Post-radiation mediastinum	4.0% (8/200)
Previous mediastinitis	0.5% (1/200)

[a] Patients may be included in more than one category
[b] Creatinine > 2.5 mg/dL, prior chest surgery, age over 75, EF < 35%

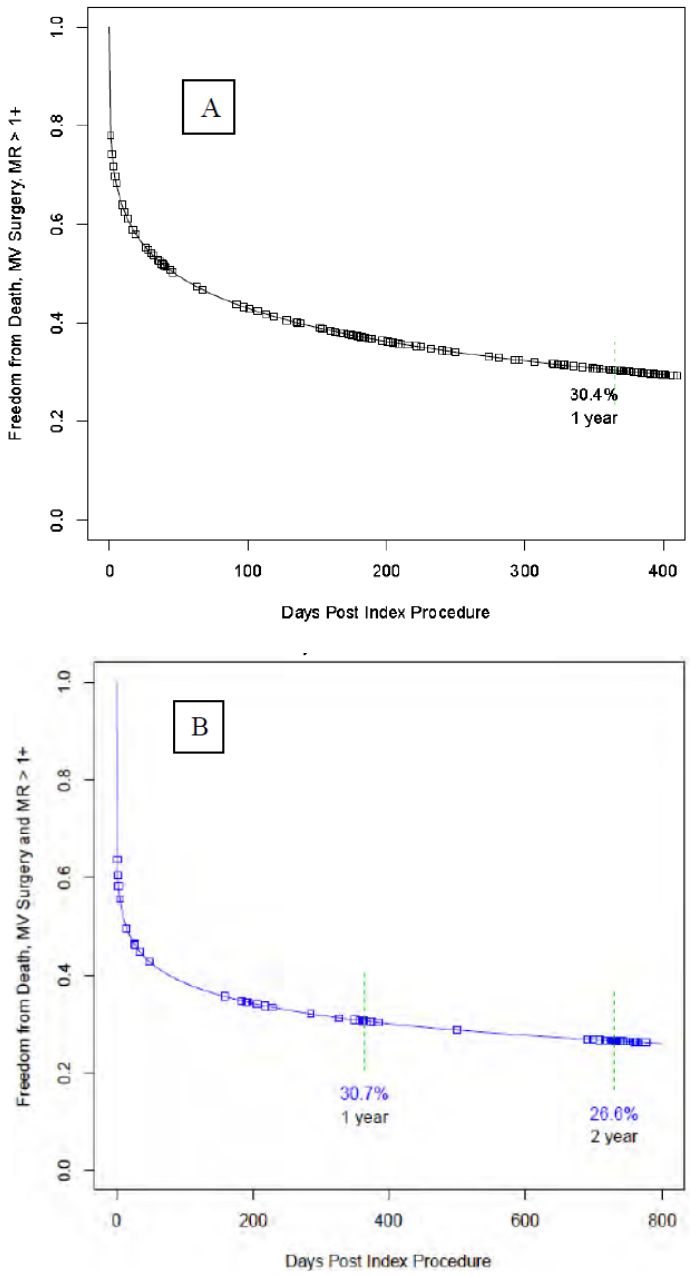
Determining the proper criteria to be used for concluding "inoperability" is critical because, as shown in the EVEREST RCT of device vs. surgery, substantial long-term effectiveness benefits are conferred following surgical relief of mitral regurgitation as compared to treatment with the MitraClip.

Figure 7: EVEREST RCT – Weibull Freedom from Death, Surgery or Reoperation, and MR > 1+ for Device and Control (Surgery) of EVEREST II RCT and Integrated High Surgical Risk Cohort.



Similar Weibull plots (Figure 8 below) for the device treated Integrated High Surgical Risk Cohort (n=351) and the REALISM Non-High Risk Cohort (n=132) reveal the same freedom from mortality, surgery or reoperation, and MR > 1+ MR at one year (30.4% and 30.7%, respectively) as seen in the device treated cohort in EVEREST RCT (30.5%).

Figure 8: Weibull Freedom from Death, surgery or reoperation, and MR > 1+, (A) = Integrated High Surgical Risk Cohort (N=351); (B) = REALISM Non-high Risk Cohort



The similarity of 1 year results in these three device treated cohorts (EVEREST RCT, Integrated High Surgical Risk Cohort and REALISM non-High Risk Cohort) despite marked differences in baseline demographics and procedural risks suggests that the observed longitudinal outcomes for the device treated cohorts may reflect the natural history of the disease in the absence of more complete initial relief of MR (i.e., surgical intervention), and that device therapy where the relief of MR is incomplete may be ineffective in altering its course.

Poolability Analysis Summary

Based on these results, and despite the presence of significant etiologic, morphologic and clinical dissimilarities sufficient to question the appropriateness of the pooling of these cohorts, patients from the two studies were considered poolable by the Sponsor and thus used to form a single Integrated High Surgical Risk cohort. As a result, the second step – propensity core matching of the Integrated High Surgical Risk Cohort to The Duke Cohort – was deemed appropriate by the sponsor in an attempt to determine whether this historical control could be used as the comparator (medical therapy) to treatment with the MitraClip device for the evaluation of safety and effectiveness of the device in high risk patients.

Pooling of the HRR and REALISM HR populations requires that reasonably similar patients are enrolled in each study. With regards to the effect these differences in etiologic and morphologic characteristics might have in judging the appropriateness of pooling, the Sponsor argues the following:

- *Despite statistical differences in LVEF and LVIDs, severe LV dysfunction was absent in both groups as a result of trial eligibility criteria. Patients enrolled in REALISM had lower LVEF and higher LVIDs than HRR patients, thus representing a higher level of surgical risk. These results were not considered clinically significant, since a since a pooled analysis represents a conservative analysis of safety for high risk patients.*
- *With respect to MR etiology, both studies enrolled patients of functional or degenerative etiologies. Subgroup analyses are provided to evaluate the impact of MR etiology on safety and effectiveness*

The Sponsor's contention that there is a higher level of surgical risk for REALISM HR patients based on lower EF and higher LVIDs is countered by objective data revealing:

- a significantly lower STS score in the REALISM HR cohort (10.5%) compared to HRR cohort (14.2%), see Table 18 above;
- a substantially lower observed 30-day procedural mortality following device treatment in the REALISM HR cohort (4.0%) compared to the HRR cohort (7.7%), see Tables 21 and 22 below;
- Lower total rates of CEC-adjudicated 30-day Major Adverse Events in the REALISM HR cohort (16.7%) versus HRR cohort (26.9%); and
- A higher rate of residual MR of 3-4+ at discharge in HRR cohort (26.3%) versus REALISM HR cohort (10.7%), excluding both dead patients and those with missing data.

Table 21: EVEREST II HRR – 30-day Procedural MFortality

	HRR (N = 78)
π_{Device}	7.7% (6/78)
95.472% UCB[a]	14.8%
$\pi_{Predicted}$	
Mean	18.2%
95% Two-sided Confidence Interval[b]	(16.4%, 20.0%)
Median	15.0%
95% Two-sided Confidence Interval[b]	(15.0%, 16.6%)

[a] UCB is based on the Clopper-Pearson method
[b] Confidence interval is calculated based on t-distribution and the confidence interval for the median is based on non-parametric methods.

Table 22: REALISM High Risk Patients (N=273) – 30-day Procedural Mortality

π_{Device}	4.0% (11/273)
95% two-sided CI[a]	(2.0%, 7.1%)
$\pi_{Predicted}$	
Mean	10.5%
95% two-sided CI[b]	(9.6%, 11.4%)
Median	8.9%
95% two-sided CI[b]	(7.9%, 10.3%)

[a] CI: Confidence Interval is based on the Clopper-Pearson method
[b] Confidence interval is calculated based on a t-distribution and the confidence interval for the median is based on non-parametric methods

When considered in total, FDA believes these factors suggest the possibility of an alternative conclusion. Specifically, the sequential and discontinuous nature of these registries resulted in experience-driven improvements in patient selection for REALISM HR, despite the presence of baseline etiologic and morphologic differences that are suggestive of a higher surgical risk cohort. The observation that more favorable acute procedural results for reduction of MR are achieved at discharge in the REALISM HR cohort versus those noted in the previously completed HRR cohort supports this notion. The sequential nature of the registries may have influenced operator experience such that selection bias (selectively enrolling patients with valvular anatomy/pathology more favorable to device therapy) or additional operator procedural experience and expertise may have significantly influenced results and rendered the groups inappropriate for pooling. Subgroup analyses to explore the cohort population differences in functional vs. degenerative MR, though hypothesis generating, may or may not be appropriate given the problems identified in pooling and the limited number of patients available for these *post-hoc* analyses performed on a limited number of propensity matched patients.

The influence of operator experience is also suggested by the similarities in sponsor defined durability of results as measured by the Primary Effectiveness Endpoint (12 month Freedom from Death, Mitral Surgery, and MR > 2+) which were similar in each cohort at 12 and 24 months (REALISM HR 63.6% and 51.7%, respectively; HRR 61.0% and 53.3%, respectively) despite a higher proportion of functional MR patients, lower mean EF, and higher LVIDs in the REALISM HR cohort – all factors that would predict a higher failure rate in similarly treated surgical patients (Alfieri stitch without a ring) over 24 months.

FDA Comment: The Integrated High Surgical Risk cohort is formed by pooling two cohorts, EVEREST II HRR and REALISM HR, in a *post hoc* manner. Neither of the cohorts represents a clinically well-defined target population, and clinically important differences

exist between them. FDA believes that the discontinuous enrollment of the two High Risk Cohorts may have been significantly impacted by selection bias considerations over time as operators gained more experience with patient selection. Given these pooling concerns, it does not seem statistically or clinically valid to use the Integrated High Surgical Risk cohort to support the new indication.

9.1.4 SAFETY AND EFFECTIVENESS

9.1.4.1 SAFETY

Following pooling of the EVEREST II HRR and REALISM HR cohorts, a comparison of observed pooled procedural mortality vs. the pooled STS score was made and is shown below (Table 23). With no control group (e.g., optimal medical management) with which to compare, the clinical meaning of both the observed device mortality and the differences between observed device mortality and the STS score are unknown. It is important to note once again that the STS score is based on the use of the STS risk calculation for MV Replacement rather than MV Repair, and that the more relevant STS score for mitral valve repair is likely to be substantially lower. It is also important to note that over half of the patients in the Integrated High Surgical Risk Cohort (178/351) were enrolled based on the use of an inflated minimal STS score. The critical shortcomings of this methodology are substantial and are also fully discussed in the High Risk Registry portion of this document.

Table 23: Integrated High Surgical Risk Cohort (N=351) – 30-day Procedural Mortality

	EVEREST Integrated High Risk Cohort (N = 351)
π_{Device}	4.8% (17/351)
95% CI[a]	(2.8%, 7.6%)
Mean	11.3%
95% CI[b]	(10.5%, 12.1%)
Median	10.3%
95% CI[b]	(9.2%, 11.5%)

[a] UCB is based on the Clopper-Pearson method

[b] Confidence interval is calculated based on two-sample t-distribution and the confidence interval for the median is based on non-parametric methods

Since no comparator was available for the pooled Integrated High Surgical Risk Cohorts, CEC-adjudicated major adverse event rates for the pooled cohort can not be appropriately compared to any other treatment in the determination of safety. It should be noted, however, that when examining the overall incidence of SAEs in the Integrated High Surgical Risk Cohort compared to the SAEs in admittedly lower risk patients undergoing surgery in the EVEREST RCT, the overall SAE rate seen after excluding transfusions is similar (9.1% Integrated High Surgical Risk Cohort vs. 11.3% in the RCT Surgical Control Cohort).

Table 24: Comparison of SAE rates – Integrated High Surgical Risk vs. RCT Surgical Control

Description of Event	Integrated High Surgical Risk Cohort (N=351)	RCT Surgery Control Patients (n=80)
Death	4.8% (17/351)	2.5% (2/80)
Myocardial infarction	1.1% (4/351)	0.0% (0/80
Re-operation for failed surgical repair or replacement	0.0% (0/351)	1.3% (1/80)
Non-elective cardiovascular surgery for adverse events	0.3% (1/351)	5.0% (4/80)
Stroke	2.6% (9/351)	2.5% (2/80)
Renal Failure	1.7% (6/351)	0.0% (0/80
Deep wound infection	0.0% (0/351)	0.0% (0/80
Ventilation > 48 hours	2.8% (10/351)	5.0% (5/80)
GI complication requiring surgery	0.3% (1/351)	0.0% (0/80
New onset of permanent AF	0.3% (1/351)	0.0% (0/80
Septicemia	0.9% (3/351)	0.0% (0/80
Transfusion ≥ 2 units	13.4% (47/351)	52.5% (42/80)
Total[a]	18.8% (66/351)	56.3% (45/80)
Total[a] (Excluding Transfusions ≥ 2 units)	9.1% (32/351)	11.3% (9/80)

FDA COMMENT: With no comparator available for analysis, it is difficult to make a safety determination for use of the MitraClip in the high risk population. In addition, it is challenging to determine the presence or absence of safety differences in high risk patient populations with severe MR, whether defined by subjective surgical risk scores, inappropriate use of STS score for MV Replacement, or by subgroups defined by specific etiologic or morphologic characteristics. Therefore, safety profile of the MitraClip relative to optimal medical management for the proposed Indication for Use and target population remains both unknown and unproven.

9.1.4.2 EFFECTIVENESS

The measures of effectiveness presented in this section include MR severity, freedom from death and MR >2+ (sponsor's definition and MR >1+ (FDA's definition), left ventricular function, NYHA Functional Class, SF-36 quality of life and re-hospitalization for heart failure.

MR severity at baseline and follow-up, for subjects who had a measurement at both baseline and follow-up visits, is presented below in Table 25. MR severity at baseline and follow-up for all patients is presented in Table 26. It is important to note that at 12 months, 36.9% (83/225) of matched cases had MR Grade ≤1+, and when considering all patients with available data, 27.8% of patients were alive with MR ≤ 1+ at 12 months.

Table 25: Integrated High Surgical Risk Cohort (N=351) – MR Grade at Baseline and Follow-up, Matched Cases[a]

MR Severity	Baseline % (n/N)	Discharge % (n/N)	Baseline % (n/N)	12 Months % (n/N)
0 : None	0.0% (0/325)	0.0% (0/325)	0.0% (0/225)	0.9% (2/225)
1+: Mild	0.6% (2/325)	46.2% (150/325)	0.4% (1/225)	36.0% (81/225)
2+: Moderate	13.2% (43/325)	39.7% (129/325)	13.8% (31/225)	46.7% (105/225)
3+: Moderate-to-severe	61.2% (199/325)	10.8% (35/325)	59.6% (134/225)	12.4% (28/225)
4+: Severe	24.9% (81/325)	3.4% (11/325)	26.2% (59/225)	4.0% (9/225)

[a] Only patients who had a measurement at both Baseline and Follow-up are included

Table 26: Accountability for MR Severity in the Integrated High Surgical Risk Cohort (N=351)

MR Severity	Baseline (n)	Discharge (n)	12 Months (n)
0	0	0	2
1+	2	157	86
2+	44	133	110
3+	208	37	28
4+	83	11	9
Missing	14	11	34
Death	0	2	82*
Total	351	351	351

% MR > 1 +** (335/337) 99%

% MR > 1 + or Dead** (183/340) 54%

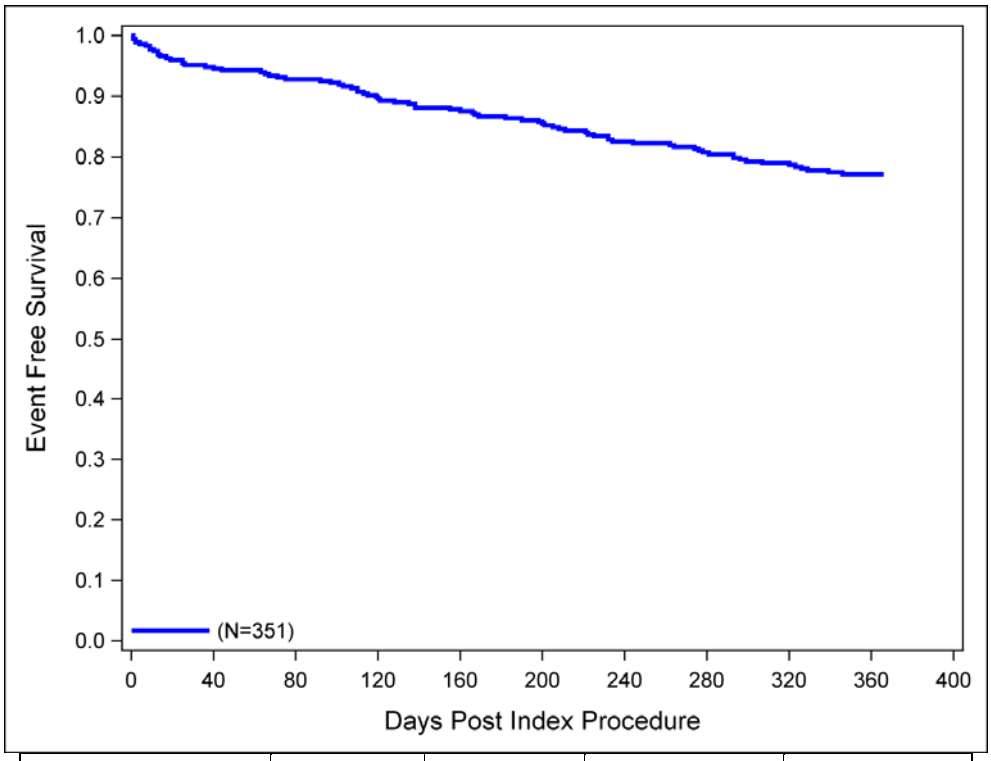
% MR > 1 + or Dead** (229/317) 72%

*This includes 4 additional deaths (deaths within 365 days post-procedure = 78; 78 + 4 = 82) that occurred between 365 and 410 days post-procedure and did not have 12 month MR assessment.
**Denominator excludes patients with missing studies (including those with surgery prior to 1 year).
Respective numbers at discharge and 12 months using MR >2+ or Death are 15% and 38%, respectively.

When considering the sponsor's definition of acute procedural success, the rate at discharge would be 85% (290/340). Using FDA's definition of acute procedural success, however, the rate at discharge is 46% (157/340).

The freedom from mortality at 12 months was 77.2% for the Integrated High Surgical Risk Cohort. Data from an appropriate control is not available for comparison, making it extremely difficult to assess whether device therapy resulted in any clinically meaningful improvement in patient outcomes.

Figure 9: Integrated High Surgical Risk Cohort (N=351) Kaplan-Meier Freedom from Mortality (deaths within 365 days post-procedure)



Time Post Index Procedure	Baseline	30 Days	6 Month	12 Month
# At Risk	351	329	295	257
# Censored	1	5	10	16
# Events	0	17	46	78
% Event Free	100%	95.1%	86.7%	**77.2%**
95% CI	-	[92.2%, 96.9%]	[82.6%, 89.8%]	[72.3%, 81.3%]

The sponsor also provides additional data meant to support a demonstration of effectiveness including sequential intra-patient measurements (baseline vs. 12 months) of selected cardiac volumes measured by echocardiograms in Integrated High Surgical Risk patients. Similar sequential measurements of LVEF were not presented. The preload and afterload states of the patients at baseline, which may significantly effect echocardiographic measures of end systolic and end diastolic volumes and dimensions, and the adequacy of their baseline medical therapy, are unknown. Since no data are available regarding the same variables at one year, comparison of data from these intervals cannot be used as definitive evidence of effectiveness.

Table 27: Integrated High Surgical Risk Cohort (N=351) – LV Function at Baseline and 12 Months, Matched Cases[a]

LV Measurement	N	Baseline	12 months	Difference (12-month - Baseline)	%Change (12-month - Baseline)
LVEDV, ml					
Mean ± SD	203	160.5 ±55.9	142.6 ±53.1	-17.9 ±31.8	-9.8 ±18.7
Median		156.3	136.3	-15.7	-9.7
97.5% UCB[b]				-13.5	-7.2
p-value (one-sided)[b]				<0.0001	< 0.0001
LVIDd, cm					
Mean ± SD	221	5.6 ±0.8	5.4 ±0.8	-0.2 ±0.4	-3.5 ±7.2
Median		5.5	5.3	-0.2	-3.1
97.5% UCB[b]				-0.1	-2.5
p-value (one-sided)[b]				<0.0001	< 0.0001
LVESV, ml					
Mean ± SD	202	87.0 ±46.8	78.9 ±43.9	-8.1 ±23.2	-6.5 ±24.3
Median		80.5	68.5	-5.9	-5.9
97.5% UCB[b]				-4.8	-3.1
p-value (one-sided)[b]				<0.0001	0.0001
LVIDs, cm					
Mean ± SD	210	4.3 ±1.1	4.1 ±1.1	-0.1 ±0.6	-1.7 ±13.9
Median		4.3	4.1	-0.0	-1.1
97.5% UCB[b]				-0.0	0.2
p-value (one-sided)[b]				0.0022	0.0430

[a] Only patients who had a measurement at both Baseline and 12 months are included
[b] 97.5% UCB and one-sided p-value are based on a t-distribution

Improvements in NYHA Functional Class and quality of life measures as assessed by the SF-36 were also noted (Table 28, below):

Table 28: Integrated High Surgical Risk Cohort (N=351) – NYHA Functional Class and Quality of Life at Baseline and 12 Months, Matched Cases[a]

Endpoint	Baseline	12 Months
NYHA Functional Class III/IV % (n/N)	82.1% (192/234)	17.1% (40/234)
Quality of Life, Physical Component Summary Score Mean ± SD (N)	34.0±9.1 (191)	38.8±11.3 (191)
Quality of Life, Mental Component Summary Score Mean ± SD (N)	44.9±13.5 (191)	49.8±12.2 (191)

[a] Only patients who had a measurement at both Baseline and 12 months are included.

Improvement in hospitalization rates were noted from the 12 months prior to enrollment to the 12 months post-discharge. It should be noted that a definition of heart failure hospitalization was not pre-specified. Furthermore, it was not clear whether patients were receiving optimal medical management during this time nor how optimal medical management was defined. These limitations make it difficult to rely heavily on this noted improvement in the evaluation of effectiveness of the device.

Table 29: Integrated High Surgical Risk Cohort (N=351) – CHF Hospitalizations

	12 months Pre-enrollment	Post-discharge through 12 Months	p-value
% Patients (n/N)	42.5% (149/351)	19.8% (67/338[a])	< 0.0001[b]
# Events	277	118	
Rate[c] (95% Two-sided Conf Int)	0.79 (0.70, 0.89)	0.41 (0.34, 0.49)	< 0.0001
# days hospitalized Mean ± SD	5.7 ± 4.7 (272)	7.1 ± 5.2 (118)	0.0061

[a] 9 patients died and 4 patients withdrew before discharge and thus do not provide data on post-discharge hospitalizations
[b] Fisher exact test
[c] p-value and confidence interval are obtained from a Poisson regression model

The Sponsor bases their conclusion that data from the Integrated High Surgical Risk Cohort supports a determination of effectiveness (clinically meaningful benefit) on the following key points:

a. KM analyses of primary effectiveness endpoint data (Sponsor preferred Freedom from Death and MR ≤ 2+) from the Integrated High Surgical Risk Cohort.

b. Intra-patient improvements in sequential echocardiographic measurements of systolic and diastolic LV dimensions (baseline vs. 12 months); and

c. Intra-patient improvements in sequential determinations of NYHA Class and SF-36 QoL measures (physical and mental components), and a decrease in the rate of re-hospitalization post-procedure.

The ability to perform a proper and complete analysis of Sponsor-submitted effectiveness outcomes by FDA, where the relative magnitude and clinical significance of observed outcomes can be appropriately judged, is challenged by:

a. FDA believes the primary effectiveness endpoint of Freedom from Death and MR ≤ 1+ is most appropriate since MR ≤ 1+ is the expected result from effective surgical repair.

b. The lack of a control population for comparison (e.g., optimal medical management) for both KM analyses conducted by the Sponsor, as well as both of the the sequential intra-patient comparisons of echocardiographic LV systolic and diastolic dimensions and measures of patient function;

c. The additional uncertainty introduced into the interpretation of intra-patient LV dimension data by the sequential nature of the echocardiographic studies and their measurements in which pre-load and afterload conditions as well as other hemodynamic variables were not controlled;

d. The lack of data for intra-patient changes observed in LVEF between baseline and 12 months; and

e. The potential for bias (placebo effect) known to effect sequential measurements in unblinded trials.

FDA Comment: The Integrated High Surgical Risk Cohort has major design limitations since it was developed by pooling two individual cohorts, each with their own weaknesses, in a *post hoc* manner. These shortcomings pose challenges to any consequential interpretation of data that would stand alone in support of a determination of safety and effectiveness, but do provide observations that are useful for hypothesis-generation necessary to guide future studies.

Recognition of these deficiencies led the Sponsor to seek to identify appropriate clinical databases that could serve as robust comparators for the Integrated High Surgical Risk Cohort. Utilizing a larger set of patient level data, a propensity matched cohort of high risk patients managed non-surgically from the Duke University Cardiac Database was obtained for comparison to a subset of the full Integrated High Surgical Risk Cohort consisting of 211 patients (HRR, n=78; REALISM HR n=133). A detailed report describing the methodology used follows in the next section.

9.2 DUKE DATABASE PROPENSITY SCORE MATCHING ANALYSIS

9.2.1 BACKGROUND

The Sponsor states that a lack of data published in the literature that could serve as a concurrent control group led them to look for an additional "real world" cohort of patients with moderate to severe MR who did not have surgery. The sponsor explored several databases, including the Duke Database for Cardiovascular Disease (DDCD) at Duke University Medical Center, which was identified to contain a patient population with MR. The Duke University Medical Center database had recorded MR severity and also had patient outcome data.

It is important to note that the sponsor did not prospectively provide the clinical protocol and detailed statistical analysis plan showing how they planned to utilize the Duke Database as a comparator to FDA for review.

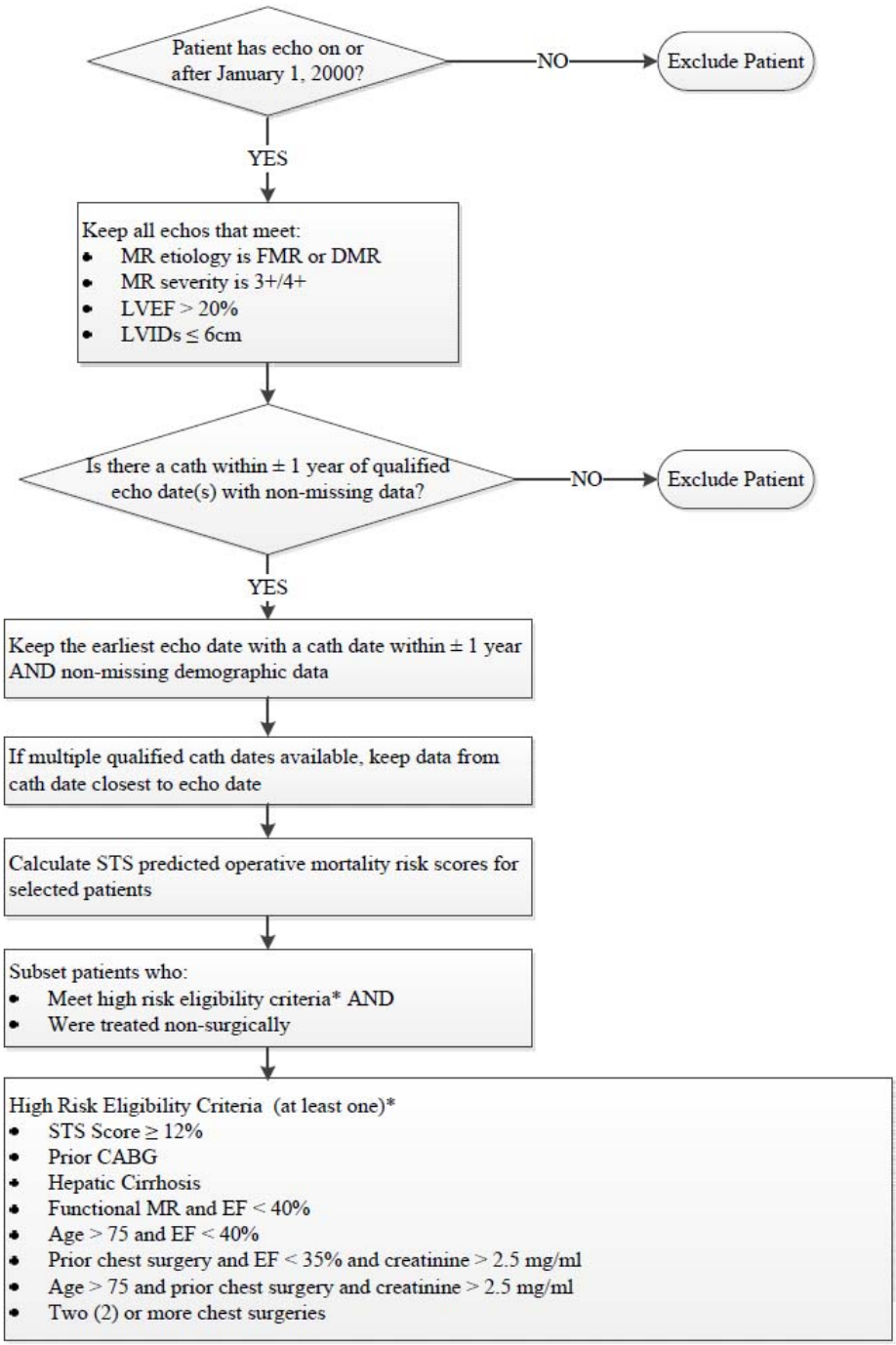
9.2.2 METHODS

9.2.2.1 Patient Selection from the Duke Databases

Patients were selected with 3+ or 4+ MR who were managed non-surgically using the echo database as the primary source. The EVEREST II High Risk study began enrollment in 2006. The Sponsor's consultants recommended to limit the Duke Database patients to the years 2000-2010. Surgical risk status of patients from Duke was determined using the definition of high surgical risk as outlined in the protocol for the EVEREST II High Risk Study. STS

scores were calculated programmatically for the patients in the Duke Cohort and validated against the STS online calculator. Therefore, the STS score was applied after the first set of patients was identified. It is not known whether a surgeon had seen any of the patients included and deemed them to be high risk for surgery. The STS scores were retrospectively calculated using data contained in the Duke database.. Figure 11 depicts the patient selection process that was used.

Figure 11: Duke Database Patient Selection Work Flow



Nine hundred and fifty three (953) patients who were managed non-surgically were identified as high surgical risk in the Duke Database (Duke Cohort) and were potential candidates for matching to the patients in the Integrated High Surgical Risk Cohort, which, at the time the Duke Analysis was conducted, contained 211 patients (MitraClip Propensity Score Analysis (PSA) Cohort) who had completed one year of follow-up.

The Statistical Analysis Plans (SAP) for the Duke Propensity Score Analysis (Duke Analysis) were not prospectively provided to FDA for review. FDA received six different versions of the Duke SAP (dated 27July2011, 10Aug2011, 09Sep2011, 14Sep2011, 21Sep2011 and 19Oct2011) after all analyses were conducted and submitted. The following is the list of issues FDA encountered upon review:

1. According to criteria specified in the Duke SAP (version 1.0, 27 July 2011), the Duke High Risk Medical Therapy Cohort (referred to as DUKE HRC: MT) of 1925 subjects was created. In the Duke SAP dated 19 Oct 2011 (Amendment 5), , it is reported that "*With agreement from all stakeholders that the level of matching achieved was adequate, analyses of mortality were performed as specified in SAP Amendment 4. However, issues were identified during validation of the programs used to create the analysis dataset containing the Duke high surgical risk patients. The dataset with 1925 records was found to contain duplicated records for some patients, thus invalidating matching and the subsequent results from the mortality analyses. In addition, issues were identified with the program used in the calculation of the STS PROM Scores. Duke and AV reviewed the programming flow used to create the analysis dataset of Duke high surgical risk patients and agreed to a simplified flow (Appendix I). The programs were modified and a new analysis dataset containing 953 records representing unique patients in the Duke Cardiac database was created.*"

   It is important to emphasize that this need for re-definition of the characteristics of the Duke High Risk patients to be used for matching was identified after outcome (mortality) analysis was conducted for matched subjects to make the characteristics more closely match the inclusion/exclusion criteria from the HRR and REALISM HR, and per the sponsor, to avoid listing of duplicate patient records. This is, concerning since appropriate use of propensity score methodology to create balance between the two treatment groups is should be conducted with outcome data concealed.

2. According to the Duke SAP (version 1.0, 27July2011), the following list of covariates were specified to be included in the logistic regression model to derive propensity scores:

Table 30: List of Covariates Specified to be Included in Logistic Regression Model

Age	Diabetes
Gender	History of renal disease
BMI	NYHA Functional Class
Previous Cardiac Surgery	History of COPD
Previous Cardiac Intervention	MR Etiology (functional or degenerative)

Previous MI	Ejection Fraction
Previous Stroke	LVIDs
History of smoking	LVIDd
History of hypertension	STS Score

Note that the highlighted variables were **not** included in the "final" logistic regression model that was used to derive propensity scores (for which results have been presented in the integrated report). It is not clear why these potentially important variables were not included.

3. In Amendment 2 (9 Sep 2011) of the Duke SAP, it is stated that *"a multivariable Cox proportional hazards model will be used to identify important predictors of mortality in the Duke High Risk Cohort."* This is another instance where outcome data (i.e., mortality data) were used even before balance between treatment groups was achieved. This may not be appropriate to implement as it may not satisfy the outcome-free requirement for propensity score analysis. This approach may not be scientifically sound and cannot be used to replace clinical judgment.

Short note on propensity score methodology and its implementation:

Propensity Score Methodology, proposed by Rosenbaum and Rubin (1983), is a method that can be used to create matched treatment-control samples or subclasses that are balanced with respect to a set of *observed* covariates, thereby potentially reducing bias or confounding due to imbalance in those covariates. Proper implementation of propensity score methodology consists of two phases: the design phase in which balance between comparison groups is sought and the analysis phase in which results regarding outcome variables are obtained. Assurance of objectivity is critical, and so the following principles need to be heeded:

1. Conceal outcome data until the design phase is complete;
2. If key covariates are not observed or very noisy, it is usually best to give up and seek a better data source ;
3. Find subgroups (subclasses or matched pairs) in which the treatment and control groups have balance – essentially, the same distribution of observed covariates – though it is not always possible to achieve balance; and
4. Protocol specified analysis should be conducted after design phase is complete.

The Duke Database was used for the years 2000-2010 using the definition of High Risk as defined for the HRR and REALISM HR. The High Risk Registry recruitment began in 2006. The Sponsor refers to the choice of comparison as a "contemporary cohort." However, there is no differentiation between those patients enrolled before 2006 and after 2006. In year 2000, medical therapy for heart failure (HF) changed with the gradual addition of beta blockers to the heart failure treatment regimen that could have altered EF, LV size, hospitalizations, and, ultimately, mortality. It is likely that more patients in 2006 were being treated by Guideline-based beta blocker recommendation than in the early 2000s.

It is also unclear how the non-eligibility for surgery was determined in the Duke Database since no information as to whether a surgeon had determined any of these patients inoperable was provided. Therefore, the reasons as to why the patients in the Duke Database did not have surgery are unclear. It should be noted that the STS calculator was used after the initial cohort was chosen. In the Duke Database, patients who did not have an outcome for any reason may have been excluded.

At the time the Duke analysis was conducted, a subset of the Integrated High Surgical Risk Cohort who had reached one year follow-up (N= 211) was used to conduct propensity score analysis (henceforth called MitraClip PSA Cohort). This included:

- 78 subjects from EVEREST II High Risk Registry (HRR); and
- 133 subjects from REALISM High Risk cohort.

.

Since the time the Duke analysis was performed, an additional 140 REALISM HR patients reached 1 year of follow-up and the Integrated High Surgical Risk Cohort analyses (except for the Duke analysis) were updated.

The following table, provided by the sponsor, compares baseline characteristics to independently assess the poolability of the 78 HRR and 133 REALISM patients.

Table 31: Poolability Analysis of the Integrated High Surgical Risk cohort used for the Duke Propensity Score Matching Analyses

Characteristic[a] % (n/N)	REALISM High Risk (N = 133)	HRR (N = 78)
Age, years		
Mean±SD (N)	75.5 ±10.6 (133)	76.7±9.8 (78)
Patients over 75 years of age	54.9% (73/133)	61.5% (48/78)
Male Gender	59.4% (79/133)	62.8% (49/78)
Body Mass Index, kg/m^2 Mean±SD (N)	26.0 ±5.9 (133)	26.6±5.0 (78)
Congestive Heart Failure	99.2% (132/133)	100.0% (78/78)
Coronary Artery Disease	79.7% (106/133)	84.2% (64/76)
Myocardial infarction	44.7% (59/132)	55.8% (43/77)
Angina	44.5% (57/128)	57.1% (44/77)
Atrial fibrillation	65.6% (80/122)	61.6% (45/73)
Cerebrovascular disease	23.3% (31/133)	17.9% (14/78)
Peripheral vascular disease	18.9% (25/132)	18.2% (14/77)
Cardiomyopathy	52.6% (70/133)	51.3% (40/78)
Hypertension	91.7% (122/133)	89.7% (70/78)
Diabetes	39.8% (53/133)	41.0% (32/78)
Moderate to Severe Renal Disease	35.3% (47/133)	23.1% (18/78)
Previous Cardiovascular Surgery, % (n/N)		
Coronary artery bypass graft	54.9% (73/133)	55.1% (43/78)
Aortic valve surgery	4.5% (6/133)	3.8% (3/78)
Tricuspid valve surgery	0.8% (1/133)	0.0% (0/78)
Other cardiac surgery	7.5% (10/133)	7.7% (6/78)
Previous PCI, % (n/N)	54.9% (73/133)	38.5% (30/78)
Cardiac Rhythm Device Implant, % (n/N)		

Characteristic[a] % (n/N)	REALISM High Risk (N = 133)	HRR (N = 78)
Pacemaker	16.2% (21/130)	22.1% (17/77)
ICD	26.2% (34/130)	13.0% (10/77)
NYHA Functional Class, % (n /N)		
I	2.3% (3/133)	0.0% (0/78)
II	14.3% (19/133)	10.3% (8/78)
III	55.6% (74/133)	60.3% (47/78)
IV	27.8% (37/133)	29.5% (23/78)
LV Ejection Fraction, % Mean±SD (N)	46.2 ±12.9(124)	54.4±13.7 (78)
LV Internal Diameter, systole, cm Mean±SD (N)	4.4±1.0.(124)	3.9±1.1 (78)
STS Predicted Mortality Risk Score[b]	11.1% ± 7.4% (133)	14.2% ± 8.2% (78)

The following table (Table 32), generated by the FDA statistician, includes additional baseline characteristics not included in the Table 31 above.

Table 32: Poolability Analysis of the Integrated High Surgical Risk cohort used for the Duke Propensity Score Matching Analyses - additional baseline characteristics not included in the Table 30 above

Characteristics	REALISM High Risk (N = 133)	HRR (N = 78)
Hypercholesterolemia	71.3% (92/129)	84.4% (65/77)
Previous Percutaneous Intervention, % (n/N)	54.9% (73/133)	38.5% (30/78)
Mitral Regurgitation Etiology, % (n /N)		
# Functional	77.4% (103/133)	59.0% (46/78)
# Degenerative	22.6% (30/133)	41.0% (32/78)

Significant differences were observed in hypercholesterolemia, previous percutaneous intervention, and cardiac rhythm device implant. Importantly, the critical etiologic (MR etiology functional vs. degenerative) and morphologic differences (LVID and LVEF) previously documented in the poolability analysis of the Integrated High Surgical Risk Cohort (n= 351: HRR, n=78; REALISM HR, n=273) persisted for this MitraClip PSA Cohort (n=211: HRR, n=78; REALISM HR, n=133).

In addition, the STS score for the MitraClip PSA Cohort is reported as 12.2% (see Table 33, below). The individual STS scores for the redefined REALISM HR cohort (n=133) and the HRR cohort prior to pooling are 11.1 ± 7.4%, and 14.2 ± 8.2%, respectively.

The demographics and baseline characteristics of the first 953 patients from the Duke Database who did not have surgery are shown in Table 33 and compared to the same characteristics of the MitraClip PSA Cohort (n=211).

Table 33: Demographic and Baseline Characteristics,
Integrated High Surgical Risk Cohort and Duke Cohort

Characteristic	MitraClip PSA Cohort (N = 211)	Duke Cohort (N = 953)
Age, years		
Mean ± SD (N)	76.0 ± 10.3 (211)	68.5 ± 13.2 (953)
Patients over 75 years of age, % (n/N)	57.3% (121/211)	36.1% (344/953)
Male Gender, % (n/N)	60.7% (128/211)	48.9% (466/953)
Body Mass Index (kg/m2)		
Mean ± SD (N)	26.2 ± 5.6 (211)	27.1 ± 6.2 (953)
Previous Cardiac Surgery, % (n/N)	58.3% (123/211)	49.9% (476/953)
Myocardial infarction, % (n/N)	48.8% (102/209)	42.8% (408/953)
NYHA III/IV, % (n/N)	85.8% (181/211)	46.6% (440/944)
COPD[a], % (n/N)	12.3% (26/211)	7.1% (68/953)
Atrial Fibrillation, % (n/N)	63.6% (124/195)	51.7% (493/953)
Stroke, % (n/N)	14.2% (30/211)	14.7% (140/953)
Diabetes, % (n/N)	40.3% (85/211)	35.5% (338/953)
Renal Disease, % (n/N)	30.8% (65/211)	18.5% (176/953)
MR Etiology, % (n/N)		
Functional	70.6% (149/211)	93.2% (888/953)
Degenerative	29.4% (62/211)	6.8% (65/953)
LV Ejection Fraction, %		
Mean ± SD (N)	49.2 ± 13.7 (201)	36.7 ± 10.9 (953)
LV Internal Diameter, systole (cm)		
Mean ± SD (N)	4.2 ± 1.1 (201)	4.2 ± 1.0 (953)
STS Predicted Operative Mortality Score	12.2 ± 7.9 (211)	9.7 ± 8.8 (953)

[a] *COPD was defined as dyspneic with the use of home* O_2

The MitraClip PSA Cohort patients are older, with over 50% >75 years old, with a higher percentage of males, as well as more comorbidities (previous surgeries, atrial fibrillation, diabetes and renal disease). The MitraClip PSA Cohort had four times as many degenerative MR patients than in the Duke Cohort but with a higher EF. The STS Score is lower in the Duke Cohort, despite the lower EF. Importantly, the MitraClip PSA Cohort had more Class III/IV symptoms in spite of significantly better EF. Therefore, at baseline, these groups are not comparable.

Figure 12: Propensity Score Distributions for MitraClip PSA Cohort (N = 211) and Duke Cohort (N = 953).

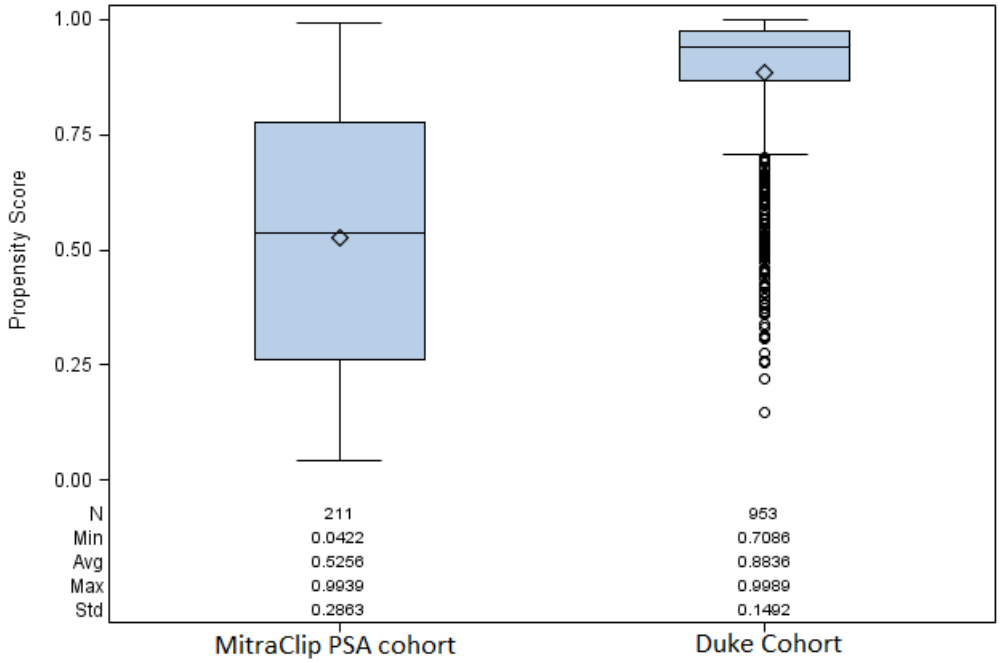


Figure 12 shows the propensity score distributions for MitraClip PSA Cohort (N = 211) and Duke Cohort (N = 953). Since the degree of imbalance between the two groups in some of the baseline and demographic characteristics such as age, LVEF and NYHA Class III/IV was large, the resulting distributions of the initial propensity scores obtained using data from all 953 patients in the Duke Cohort did not yield sufficient overlap. Therefore, data from the Duke Cohort were "trimmed" as follows: patients in the Duke cohort with age and/or LVEF outside the range of the corresponding mean ± 1.5 SD from the MitraClip PSA Cohort were excluded from further consideration. Thus, patients in the Duke Cohort with continuous variable data who were unlikely to be "good matches" to the MitraClip PSA Cohort patients were excluded, resulting in a "trimmed" Duke Cohort (N = 527). Matches for the MitraClip PSA Cohort were then obtained from the Trimmed Duke Cohort.

9.2.2.2 Trimmed Duke Cohort

The Trimmed Duke Cohort consists of 527 patients shown in Table 34 below.

Table 34: Demographic and Baseline Characteristics,
EVEREST High Risk and "Trimmed" Duke Cohorts

Characteristic	MitraClip PSA Cohort (N = 211)	Trimmed Duke Cohort (N = 527)
Age, years		
Mean ± SD (N)	76.0 ± 10.3 (211)	74.7 ± 7.9 (527)
Patients over 75 years of age, % (n/N)	57.3% (121/211)	50.1% (264/527)
Male Gender, % (n/N)	60.7% (128/211)	46.3% (244/527)
Body Mass Index (kg/m2)		
Mean ± SD (N)	26.2 ± 5.6 (211)	26.9 ± 5.8 (527)
Previous Cardiac Surgery, % (n/N)	58.3% (123/211)	60.2% (317/527)
Myocardial infarction, % (n/N)	48.8% (102/209)	44.4% (234/527)
NYHA III/IV, % (n/N)	85.8% (181/211)	40.2% (209/520)
COPD[a], % (n/N)	12.3% (26/211)	7.2% (38/527)
Atrial Fibrillation, % (n/N)	63.6% (124/195)	57.9% (305/527)
Stroke, % (n/N)	14.2% (30/211)	15.6% (82/527)
Diabetes, % (n/N)	40.3% (85/211)	40.0% (211/527)
Renal Disease, % (n/N)	30.8% (65/211)	20.5% (108/527)
MR Etiology, % (n/N)		
Functional	70.6% (149/211)	90.5% (477/527)
Degenerative	29.4% (62/211)	9.5% (50/527)
LV Ejection Fraction, %		
Mean ± SD (N)	49.2 ± 13.7 (201)	41.7 ± 9.6 (527)
LV Internal Diameter, systole (cm)		
Mean ± SD (N)	4.2 ± 1.1 (201)	3.9 ± 0.9 (527)
STS Predicted Operative Mortality Score	12.2 ± 7.9 (211)	11.5 ± 8.9 (527)

[a] *COPD was defined as dyspneic with the use of home* O_2

The NYHA Class is still significantly higher in the MitraClip PSA Cohort than in the Trimmed Duke Cohort. In addition, the rates of atrial fibrillation, male gender, COPD, and renal disease were higher in the MitraClip PSA Cohort. The Trimmed Duke Cohort had a higher incidence of functional MR, a lower mean EF but had smaller ventricles overall (LVIDs).

Three levels of Propensity score matching were performed.

Matched Cohort 1: A caliper size of a quarter (0.25) of the (average) standard deviation of the logit of the propensity scores was used. Patients were first matched to within a caliper of 0.25. If multiple matches were identified for a MitraClip PSA Cohort, the match was narrowed to a single patient based on the smallest Mahalanobis distance.

Matched Cohort 2: The caliper size was expanded to 0.4 from 0.25 for patients with no matches within the narrower caliper.

Matched Cohort 3: Finally, MitraClip PSA Cohort patients with no matches within the expanded caliper of 0.4 were matched to patients in the Duke High Risk Cohort with the nearest propensity score.

The following variables were used in the logistic regression model: age, gender, previous MI, previous stroke, COPD, history of renal disease, diabetes, previous cardiac surgery, NYHA

Functional Class III/IV, and LVEF. The Sponsor states that MR etiology was not included in the model because other baseline co-morbidities such as previous MI, lower LVEF, and previous cardiac surgery are highly correlated with functional etiology.

FDA Comment: The sponsor did not include all clinically relevant variables such as MR etiology and ventricular size in this model. Functional MR can occur in non-ischemic HF as well and adds a high risk factor for outcomes in addition to EF. The FDA believes that EF, and functional MR, although related, contribute independently to outcomes and neither should be excluded from the model.

9.2.2.2.1 Matched Cohort 1

Matched Cohort 1 was derived using the narrowest caliper size of 0.25*SD of the logit of the propensity score. One hundred and twenty seven (127) of the 211 MitraClip PSA Cohort patients were matched 1:1 to patients in the Trimmed Duke Cohort. The Matched Cohort 1 subset of patients who received the MitraClip have larger ventricles, higher incidence of stroke and diabetes, and a higher proportion of degenerative etiology, but a lower STS score. The timeline during which these Duke control patients had an echo and were seen is unclear. Left ventricular size is related to mortality, although the mean LVID is relatively small. It is unknown whether patients in the Duke Database received optimal medical management by standards contemporaneous with the MitraClip groups (after 2008). There are no data provided about background medical therapy for patients in the Duke Database, and the rates of mortality may differ by the years with better outcome after 2003 or 2004 when beta blockers were included in the Guidelines. No data were provided on the QRS width or use of CRT in Duke patients in the years after the MIRACLE trial results were published, which is important since CRT can markedly improve MR.

Table 35: Demographic and Baseline Characteristics – Matched Cohort 1 (Caliper size = 0.25)

Characteristic	Cohort 1 MitraClip Patients (N = 127)	Cohort 1 Duke Patients (N = 127)	p-value
Age, years			
Mean ± SD (N)	74.6 ± 10.5 (127)	74.6 ± 7.9 (127)	0.547
Patients over 75 years of age, % (n/N)	52.8% (67/127)	51.2% (65/127)	0.802
Male Gender, % (n/N)	52.8% (67/127)	51.2% (65/127)	0.802
Body Mass Index (kg/m2)			
Mean ± SD (N)	26.2 ± 4.8 (127)	27.2 ± 5.2(127)	0.093
Previous Cardiac Surgery, % (n/N)	55.1% (70/127)	52.8% (67/127)	0.706
Myocardial infarction, % (n/N)	44.4% (56/126)	44.9% (57/127)	0.944
NYHA III/IV, % (n/N)	78.0% (99/127)	74.8% (95/127)	0.555
COPD[a], % (n/N)	9.4% (12/127)	5.5% (7/127)	0.233
Atrial Fibrillation, % (n/N)	56.5% (65/115)	61.4% (78/127)	0.439
Stroke, % (n/N)	14.2% (18/127)	8.7% (11/127)	0.167
Diabetes, % (n/N)	44.9% (57/127)	37.0% (47/127)	0.202
Renal Disease, % (n/N)	28.3% (36/127)	26.0% (33/127)	0.672
MR Etiology, % (n/N)			
Functional	84.3% (107/127)	88.2% (112/127)	0.363
Degenerative	15.7% (20/127)	11.8% (15/127)	0.363
LV Ejection Fraction, %			
Mean ± SD (N)	43.0 ± 11.8 (120)	44.1 ± 9.8 (127)	0.351
LV Internal Diameter, systole (cm)			
Mean ± SD (N)	4.5 ± 1.0 (121)	3.9 ± 1.0 (127)	< 0.0001
STS Predicted Operative Mortality Score	11.1 ± 7.1(127)	13.2 ± 10.7(127)	0.322

[a] *COPD was defined as dyspneic with the use of home O_2*

The propensity score matching algorithm used did not provide a match for all 211 Integrated High Surgical Risk Cohort subjects. As a result, the Matched Cohort 1, contains only a subset (n=127) of the 211 MitraClip PSA cohort patients, which itself is a subset of the Integrated High Surgical Risk Cohort (n=351) comprised of two sequentially enrolled pooled cohorts (HRR, n=78; REALISM HR, n=273) with significant differences in critical baseline etiologic, morphologic and procedural risk variables. This is of concern to FDA as the ultimate make-up of these 127 matched subjects is unknown and does not represent any well-defined population.

9.2.2.2.2 Matched Cohort 2

Matched Cohort 2 was obtained using an expanded caliper size of 0.4*SD of the logit of the propensity score. This cohort contained all patients who were matched within a caliper of 0.25*SD of the logit of the propensity score, in addition to patients who are matched with a slightly larger caliper size of 0.4*SD of the logit of the propensity score. Five additional patient matches were obtained with the expanded caliper width.

The sponsor reports that the results for Matched Cohort 2 are similar to Matched Cohort 1 including the differences in ventricular size, stroke, diabetes and degenerative etiology. It

should be noted that the Matched Cohort 2 MitraClip patients also had slightly more NYHA Class III/IV patients.

Once again, the make-up of these 132 matched subjects is of concern to FDA because these patients do not represent any well-defined population.

9.2.2.2.3 Matched Cohort 3

To generate Matched Cohort 3, patients from MitraClip PSA Cohort with no matches within the expanded caliper of 0.4 were matched to patients in the Trimmed Duke Cohort with the nearest propensity score. Though this results in including all 211 EVEREST MitraClip PSA Cohort subjects, Matched Cohort 3 MitraClip patients included a higher proportion of men, more NYHA Class (III/IV), higher incidence of COPD, renal disease, higher proportion of degenerative etiology, higher EF, but larger ventricles and similar STS score. It is not clear how the STS score can be similar with important mismatches in NYHA and EF.

The distribution of propensity scores for Matched Cohort 3 subjects does not provide sufficient overlap and thus does not provide balance across important clinical covariates.

9.2.2.2.4 Functional Mitral Regurgitation (FMR) subset of Cohort 1

Further subsetting of Matched Cohort 1 to include only patients with functional MR (ischemic etiology unclear) results in a subset showing that MitraClip patients are more symptomatic, have larger ventricles, lower EF, higher incidence of stroke, diabetes and renal disease, but a lower STS score. The subsetting of Matched Cohort 1 to form the FMR subset of Matched Cohort 1 confounds the issues identified with Match Cohort 1 above. FDA believes that the FMR subset does not represent a well-defined population.

Table 36: Demographic and Baseline Characteristics
FMR subgroup from Matched Cohort 1 (Caliper Size 0.25)

Characteristic	Cohort 1 MitraClip FMR patients (N = 107)	Cohort 1 Duke FMR Patients (N = 112)	p-value
Age, years			
Mean ± SD (N)	73.0±10.4 (107)	73.9±7.6 (112)	0.844
Patients over 75 years of age, % (n/N)	45.8% (49/107)	48.2% (54/112)	0.720
Male Gender, % (n/N)	52.3% (56/107)	52.7% (59/112)	0.960
Body Mass Index (kg/m2)			
Mean ± SD (N)	26.1±4.9 (107)	27.3±5.1(112)	0.057
Previous Cardiac Surgery, % (n/N)	55.1% (59/107)	53.6% (60/112)	0.816
Myocardial infarction, % (n/N)	49.1% (52/106)	47.3% (53/112)	0.798
NYHA III/IV, % (n/N)	81.3% (87/107)	73.2% (82/112)	0.154
COPD[a], % (n/N)	9.3% (10/107)	6.3% (7/112)	0.392
Atrial Fibrillation, % (n/N)	56.7% (55/97)	60.7% (68/112)	0.557
Stroke, % (n/N)	15.9% (17/107)	8.9% (10/112)	0.117
Diabetes, % (n/N)	46.7% (50/107)	35.7% (40/112)	0.098
Renal Disease, % (n/N)	30.8% (33/107)	26.8% (30/112)	0.508
LV Ejection Fraction, %			
Mean ± SD (N)	40.7±10.0 (103)	43.0±9.6 (112)	0.144
LV Internal Diameter, systole (cm)			
Mean ± SD (N)	4.7±0.9 (104)	4.0±0.9 (112)	< 0.0001
STS Predicted Operative Mortality Score	10.9±7.3(107)	12.9±11.1(112)	0.526

[a] *COPD was defined as dyspneic with the use of home O_2*

9.2.2.2.5 Unmatched MitraClip patients from Cohort 1

After propensity score matching was performed in Cohort 1, 40% (84/211) of the MitraClip PSA Cohort remained unmatched. Characteristics of the matched and unmatched MitraClip patients in Cohort 1 show that the unmatched patients are older, include a higher proportion of men, more previous surgery, more NYHA Class III-IV, higher incidence of COPD, renal disease, degenerative etiology, higher LV EF and smaller ventricles, with much higher STS score.

Table 37: Demographic and Baseline Characteristics
Matched Cohort 1 and Unmatched MitraClip patients in Cohort 1

Characteristic	Cohort 1 Matched MitraClip Patients (N = 127)	Cohort 1 Unmatched MitraClip Patients (N = 84)
Age, years		
Mean ± SD (N)	74.6 ± 10.5 (127)	78.0 ± 9.7 (84)
Patients over 75 years of age, % (n/N)	52.8% (67/127)	64.3% (54/84)
Male Gender, % (n/N)	52.8% (67/127)	72.6% (61/84)
Body Mass Index (kg/m2)		
Mean ± SD (N)	26.2 ± 4.8 (127)	26.4 ± 6.7(84)
Previous Cardiac Surgery, % (n/N)	55.1% (70/127)	63.1% (53/84)
Myocardial infarction, % (n/N)	44.4% (56/126)	55.4% (46/83)
NYHA III/IV, % (n/N)	78.0% (99/127)	97.6% (82/84)
COPD[a], % (n/N)	9.4% (12/127)	16.7% (14/84)
Atrial Fibrillation, % (n/N)	56.5% (65/115)	73.8% (59/80)
Stroke, % (n/N)	14.2% (18/127)	14.3% (12/84)
Diabetes, % (n/N)	44.9% (57/127)	33.3% (28/84)
Renal Disease, % (n/N)	28.3% (36/127)	34.5% (29/84)
MR Etiology, % (n/N)		
Functional	84.3% (107/127)	50.0% (42/84)
Degenerative	15.7% (20/127)	50.0% (42/84)
LV Ejection Fraction, %		
Mean ± SD (N)	43.0 ± 11.8 (120)	58.4 ± 11.0 (81)
LV Internal Diameter, systole (cm)		
Mean ± SD (N)	4.5 ± 1.0 (121)	3.9 ± 1.0 (80)
STS Predicted Operative Mortality Score	11.1 ± 7.1(127)	14.0 ± 8.6(84)

[a] *COPD was defined as dyspneic with the use of home O_2*

9.2.3 RESULTS

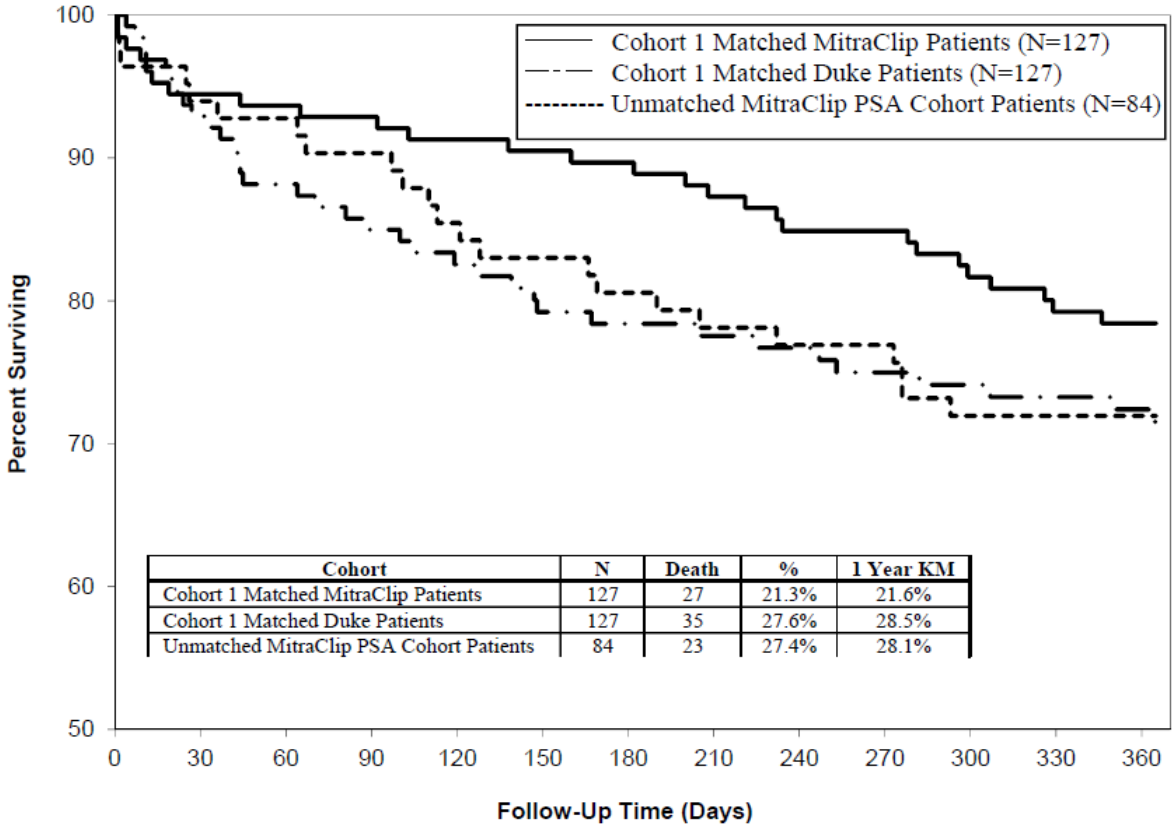
Given all the concerns outlined above, the interpretation of the following results extracted from sponsor's report is not clear.

9.2.3.1 Matched Cohort 1 – All-Cause Mortality

Figure 13 below presents the Kaplan-Meier freedom from all-cause mortality curve for Matched Cohort 1 (127 matched patients from MitraClip PSA Cohort) compared to the matched Duke High Risk cohort (N=127) and the remaining unmatched EVEREST HR data from Cohort 1 (N=84).

For 953 Duke Cohort patients, Duke did not provide FDA the Echo date, Death date and other follow-up visit dates to calculate the duration of follow-up used in K-M plot although they did provide durations between events since Duke believed the exact dates were a patient confidentiality issue. Thus, FDA was unable to reproduce the duration of follow-up for each Duke subject used in generating the K-M plot.

Figure 13: Kaplan-Meier Freedom from All-Cause Mortality
Unmatched EVEREST High Risk Patients vs. Matched Cohort 1



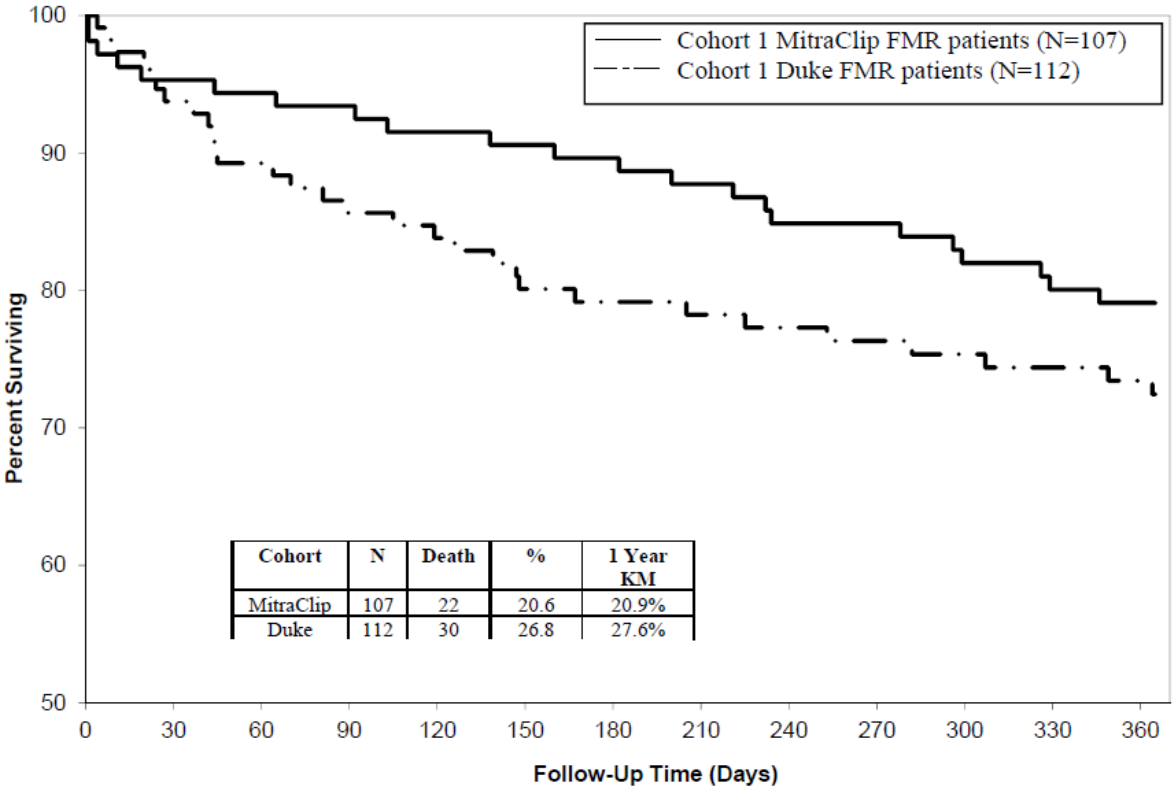
Cohort	N	Death	%	1 Year KM
Cohort 1 Matched MitraClip Patients	127	27	21.3%	21.6%
Cohort 1 Matched Duke Patients	127	35	27.6%	28.5%
Unmatched MitraClip PSA Cohort Patients	84	23	27.4%	28.1%

Early mortality is higher in the unmatched MitraClip PSA Cohort than in the non-surgical Matched Duke patient group but the curves approximate each other after the first 3 months. This KM graph is difficult to interpret, particularly since the highest risk group that is unmatched has STS scores that more closely resemble those for the Matched Duke Cohort 1 patients and exhibit a similar mortality rate.

9.2.3.2 FMR Subset of Matched Cohort 1 – All-Cause Mortality

Figure 14 below presents the Kaplan-Meier freedom from all-cause mortality curve for the subset of Matched Cohort 1 MitraClip patients with functional MR (FMR subset of Matched Cohort 1, N=107) compared to the subset of Matched Cohort 1 Duke patients with FMR (N=112).

Figure 14: Kaplan-Meier Freedom from All-Cause Mortality
FMR Subgroup of Matched Cohort 1



This KM graph is difficult to interpret, particularly in light of the difficulties of interpreting the Matched Cohort 1 K-M plot, of which this is a subset.

9.2.4 Conclusions Based on Duke Analysis

In conclusion, the patients in the Duke databases were accumulated in an era where significant changes in medical and device therapy of CHF have occurred. We have no data on either the type or adequacy of medical or device therapy used to treat the registry patients. Duke was an EVEREST II and REALISM investigational site and there was no information provided on whether any of the Duke patients in the database had been offered to enter the MitraClip registry or, conversely, if any patients who might have been captured in the Duke databases were consequently not captured because the patients were enrolled in the trials. During the years included in the Duke cohort there were 4 patients who were enrolled in the in either the HRR or REALISM HR studies, but it is not know how many patients were not enrolled in these studies who would have qualified.

FDA Comment: The Duke Propensity Score Analysis was a retrospective, subset analysis with results that are difficult to interpret and where the matched cohorts do not represent any well-defined population. Given all the concerns regarding the creation of the matched cohorts above, the difficulties in interpreting the results, and considering the *post-hoc* nature of this

analysis in a cohort that is not well-defined through different time periods, this analysis should be viewed with extreme caution and should be considered hypothesis generating.

9.3 OHIO STATE DATABASE COMPARISON

9.3.1 BACKGROUND

The Sponsor identified the patient database of the Ohio State University Medical Center: Heart and Vascular Center (OSU) to contain a patient population that could serve as a comparator for EVEREST High Risk patients.

9.3.2 METHODS

Access was gained to the Ohio State University patient database containing data on patients with a diagnosis of MR by diagnostic codes (ICD) between January 1, 2006 and December 31, 2010, who were managed non-surgically. Specifically, MR patients with any ICD code of surgery or intervention were excluded.

The high risk criteria from EVEREST II High Risk Registry were applied to the existing OSU data. The resulting demographics of the OSU Cohort of 181 patients are presented below in Table 38. It is worth noting that the data from the OSU Cohort were collected over a different time period (2006-2010) as compared to the Duke Database. As noted above, these are different eras with different medical and device therapies available.

Table 38: Demographic and Baseline Characteristics

Characteristic	High Risk Historical OSU Cohort (N=181)
Age, Years Mean±SD (N) Patients over 75 years of age, % (n/N)	 78.4±4.6 (181) 75.7% (137/181)
Male Gender, % (n/N)	65.7% (119/181)
Previous Cardiac Surgery, % (n/N)	24.3% (44/181)
Chronic Lung Disease, % (n/N)	33.7% (61/181)
Peripheral Vascular Disease, % (n/N)	28.7% (52/181)
Cerebrovascular Disease, % (n/N)	2.8% (5/181)
History of Dialysis	12.2% (22/181)
LV Ejection Fraction, % Mean±SD (N)	 35.9±15.4 (73)

9.3.3 RESULTS

The all-cause hospitalization rate of the OSU Cohort (N=181) was 0.43 per 12 months. The Integrated High Surgical Risk Cohort had a higher rate of hospitalization prior to MitraClip at 0.79 per 12 months. The Freedom from Mortality in the OSU Cohort is presented below in Table 39. The 12-month mortality reported is lower (18%) compared to that reported in the

Duke database (26%). There are large amounts of missing data for the OSU cohort, including those of comorbidities.

Table 39: Kaplan-Meier Estimates of Freedom from All-Cause Mortality

Time After MR Diagnosis	30 Days	12 Months	24 Months
% Event Free	94.5%	82.1%	67.6%
95% CI	(90.0%, 97.0%)	(75.4%, 87.1%)	(59.3%, 74.5%)

FDA Comment: When considering the missing data, including the lack of MR severity, NYHA Class, and LV measurements, the OSU database provides results that are difficult to compare to the Integrated High Surgical Risk Cohort, providing little added value to the review of the MitraClip CDS.

10 POST APPROVAL STUDY

Note: The inclusion of a Post-Approval Study section in this summary should not be interpreted to mean that FDA has made a decision or is making a recommendation on the approvability of this PMA device. The presence of a post-approval study plan or commitment does not in any way alter the requirements for premarket approval and a recommendation from the Panel on whether the risks outweigh the benefits. The premarket data must reach the threshold for providing reasonable assurance of safety and effectiveness before the device can be found approvable and any post-approval study could be considered. The issues noted below are FDA's comments regarding potential post-approval studies, for the Panel to include in the deliberations, should FDA find the device approvable based upon the clinical premarket data.

If the MitraClip is approved, the FDA has identified the following postmarket concerns, through review of the premarket data, and recommends that a post-approval study (PAS) should be conducted for the following reasons:

1. Assess longer-term performance of the device – As an innovative device, understanding of the longer-term safety and effectiveness from implant to death or surgery (if considered) for the approved indication is necessary.

2. Determine device performance in the broader population – The expected community of users and patients may differ from the premarket and this may impact device performance.

3. Assess the effectiveness of training programs – Those who have performed the procedure were selected based on certain criteria and have received extensive training. It is unclear how less trained and experienced operators will fair when implanting the device.

4. Evaluate device performance in patient sub-groups – Long-term effectiveness of the device among patients with mitral regurgitation of different etiology or levels of surgical risk factors has not been established in the US.

5. Monitor rare adverse events and real world experience – As a first of a kind device, the postmarket performance of the device should be assessed for the occurrence of rare or unexpected adverse events that may not have been detected within the premarket study.

The sponsor submitted the PAS protocol proposal (Rev. 001) on August 28, 2012. An overview of the proposed PAS protocol is provided below. Concerns about the PAS protocol proposal are included in the assessment following the proposal overview.

Overview of Proposed Post-Approval Study

Post-Approval Study Objectives

1. Confirmation of the long-term safety and effectiveness of the MitraClip device in the commercial use setting;

2. Confirmation that the MitraClip device can be used safely by implanting physicians with different levels of experience; and

3. Identification of any low-frequency or unanticipated MitraClip device-related adverse events that may occur in a commercial setting.

Study Design and Study Population

This is a prospective, non-randomized, single arm multi-center study evaluating the MitraClip for the treatment of moderate-to-severe or severe MR in patients deemed too high risk to undergo mitral valve surgery. There are two separate arms for the study comprising of Functional mitral regurgitation (FMR) patients and Degenerative mitral regurgitation (DMR) patients.

Hypotheses

The primary safety hypothesis is that freedom from the composite endpoint of death and stroke at 12 months is greater than the performance goal of 70%. The primary safety hypothesis was based on the event rates observed in 351 Integrated High Surgical Risk patients who completed 1 year of follow-up.

The primary effectiveness hypothesis is that freedom from mitral valve surgery at 24 months is greater than the performance goal of 80%. The primary effectiveness hypothesis was based on 24-month REALISM High Risk data for the freedom from mitral valve surgery.

The secondary safety hypothesis is that the proportion of subjects free from composite of safety events of all-cause death, stroke (major and minor), MI, and non-elective cardiovascular surgery for device-related complications at 30 days is greater than 80%.

Endpoints

Primary Safety Endpoint

- A composite of death and stroke through 12 months

Primary Effectiveness Endpoint
- Mitral valve surgery through 24 months

Secondary Safety Endpoints
- A composite of all-cause death, stroke (major and minor), myocardial infarction, and non-elective cardiovascular surgery for device-related complications in the device group at 30 days

Additional Endpoints
1. Any device-related adverse events including single leaflet device attachment, MitraClip device, or MitraClip component embolization
2. Device- and procedure-related endpoints – implant rate, device procedure time, total procedure time, and device time
3. Clinical Endpoints

endpoints	baseline	discharge	30 days	12 mos	24 mos	annually thereafter through 5 years
Proportion of surviving patients with reduction in MR Severity to < 2				x	x	
Proportion of patients with improved Six minute Walk Distance>24 m	x			x	x	
Proportion of patients with reduction in NYHA to class II	x			x	x	
Annual rate of freedom from mitral valve surgery and death				x	x	x
Annualized heart failure hospitalization res				x	x	
Surgical mitral valve repair rate(stratified by MR etiology)				x	x	x
Left ventricular end systolic volume (LVESV)	x			x	x	
Left ventricular internal dimension at end systole (LVIDs)	x			x	x	
Proportion of surviving patients with reduction in MR Severity to < 1+				x	x	
Proportion of patients with improved KCCQ Quality of Life Score >5 pts	x			x	x	
Annual rate of freedom from death (all cause, cardiovascular cause)				x	x	x
All new onset atrial fibrillation			x	x	x	x
Surgical mitral valve replacement rates (stratified by MR etiology)				x	x	x
Left ventricular end diastolic volume (LVEDV)	x			x	x	
Left ventricular internal dimension at end diastole (LVIDd)	x			x	x	

Enrollment Plan and Follow-up

The PAS will consecutively enroll patients who are similar to patients enrolled in the EVEREST II High Risk Registry and REALISM High Risk studies. For a patient to be enrolled, 2 cardiothoracic surgeons, one with at least 25 mitral valve repair experience in the prior year, must agree that the patient is too high risk for mitral valve surgery.

Patients who receive the MitraClip will be followed through 5 years. The protocol does not state any measures for preventing and controlling loss to follow-up.

Sample size

A total of 2400 patients will be enrolled in the study, with 1200 patients in each study arm (FMR arm and DMR arm). The sample size and power for the primary safety endpoint or primary effectiveness endpoint were calculated by performing 1000 simulations; these analyses indicate that a sample of 1200 patients in each arm, accounting for 10% attrition at 12 months, provides at least 90% power to reject the null hypothesis at the 5% significance level.

Statistical Plan

All endpoints will be analyzed separately for the DMR and FMR populations. The test of the null hypothesis for the primary endpoints (safety and effectiveness) will be the Kaplan-Meier survival estimate, and together with the variance estimated by the Greenwood Method, may be used to set up the test of the null hypothesis as a Z-test. The null hypothesis will be rejected at the 5% significance level if the test statistic is less than -1.645. For the composite 30-day secondary endpoint an exact test for a single proportion will be performed at the 5% level of significance.

All PAS outcomes will be analyzed by both physician level of experience and site level of experience to ensure adequate effective training. The evaluation metrics for the training program to be provided include:

1. Procedural Safety Metrics: Device time, implant rate, and reduction of MR; and

2. Post-Procedural Safety Metrics: Number of single leaflet device attachments (SLDA) and number of device embolizations.

Additional criteria have been proposed for monitoring of adverse events through the use of the cumulative sum control charts at each site, with details for analyzing the charts.

Timeline

The timeline for the study implementation was not provided in the submission.

FDA Comments on Proposed Post-Approval Study

Study Objectives

- The current PAS protocol proposal intended to confirm longer-term safety and effectiveness of the MitraClip Device would evaluate the primary safety and primary effectiveness endpoints at 12 months and 24 months post–intervention, respectively, and a composite secondary safety endpoint at 30 days. There is the need to provide

primary endpoint assessment at timepoints beyond 24 months in order to reflect longer-term performance of the device.

- There is a proposal to identify low-frequency or unanticipated device-related adverse events that may occur, but these events are not specified and there are no proposals as to how the events will be monitored, captured or evaluated.

Study Design

- The proposal to evaluate primary effectiveness by assessing freedom from mitral valve surgery as a surrogate for device effectiveness is considered unacceptable for the following reasons:

    - The patients are deemed too high surgical risk or "inoperable," therefore, even if the device totally fails, the patient may not be offered mitral valve surgery.

    - There are several factors that determine if a patient would have mitral valve surgery in the event that the device fails (i.e., patient consenting to surgery, surgery being offered to the patient, the patient's other co-morbid conditions such as renal failure, COPD, etc.).

    - The effectiveness endpoint is now mitral valve surgery alone as opposed to mitral valve surgery and death, but the protocol states that data will be collected to assess the "annual rate of freedom from mitral valve surgery and death."

  For this PAS, FDA proposes a combination of using medical management as a concurrent control and a surgical control to effectively characterize performance of the MitraClip in the postmarket setting.

- Using results from the sponsor's previous study on the same device as a comparator is not appropriate for the postmarket evaluation of device performance. In "real world" practice, it is unclear whether providers and patients will be deciding between the device and surgery or between the device and continued medical management. FDA recommends inclusion of both surgery and continued medical management as concurrent comparison arms to effectively characterize "real world" performance. Concurrent study arms (device; surgery; and medical management) should be included in the PAS protocol.

- The sponsor will work in collaboration with the Society of Thoracic Surgeons (STS) and American College of Cardiology (ACC) societies to conduct this post-approval study. As part of this collaboration, it is recommended that the data collection for this study (i.e., pre-procedure, peri-procedure, post-procedure, discharge, 30-day, and 1-year follow-up) be nested within the National Transcatheter Aortic Valve Replacement (TVT) registry. This would be done as a registry module focused on transcatheter mitral valve therapies. Within the TVT registry, patients implanted with MitraClip for the intended indication could also be followed for a longer period

of time, 5-10 years or more, which could serve as a repository data source for device performance evaluation over a much longer time than the PAS.

Sample size

- To confirm the sample size calculations, the test statistic used needs to be clearly specified, and the simulation program code needs to be provided as well.

- The protocol describes the process and strategy for clinical site selection, but does not specify the minimum number of sites, or the number of new and old sites.

- The rates of death and stroke for both DMR and FMR are approximately 27%, therefore, it is acceptable that the power analyses did not vary by disease and the study is powered per arm.

Outcomes

- The independent Clinical Events Committee (CEC) will be responsible for adjudicating primary and secondary safety endpoints based on specific definitions of clinical events in the trial. This is adequate.

- Specific definitions are not provided for major adverse events (MAE; e.g., renal failure, upper-GI bleeding complications, myocardial infarction, etc.). The primary effectiveness endpoint for the PAS differs from the premarket endpoint; it does not include evaluation of MR (effectiveness = MR $\leq$ 2+). The sponsor must provide definitions of all endpoints and justify any differences in the premarket and postmarket endpoint safety definitions.

Enrollment and follow-up

- The minimum number of study sites required to enroll study subjects and the maximum number of subjects required per site were not stated in the protocol proposal

- The proposal to enroll 2400 new patients who are similar to patients enrolled in patients enrolled in the EVEREST II High Risk Registry and REALISM High Risk to be implanted with the MitraClip. However, inclusion/exclusion criteria for patient selection were not included in the protocol.

- No justification is provided for 5 years of follow-up to characterized long term performance. To assess relevant durability and long-term clinical outcomes in patients with the device, the FDA recommends a minimum 10 year follow-up.

- The experience levels of the device users are adequately defined. Level 1 is defined as an Operator with no prior experience implanting the device or with no current

experience; and Level 2 is defined as an Operator who has prior experience implanting the device.

- Loss to follow-up can introduce study bias. Detailed measures that will prevent and handle loss to follow-up need be included in the protocol.

Statistical plan

- For primary endpoint analyses, the proposal to use Kaplan-Meier survival estimate, together with the variance estimated by the Greenwood method, to set up the test of the null hypothesis as a Z-test looks appropriate. The proposal does not detail how missing covariates or outcome data will be addressed in the analyses.

11 FDA SUMMARY OF ISSUES

In this section, we have summarized the key FDA findings based on review of the clinical data presented by the sponsor.

The EVEREST II RCT did not demonstrate an appropriate benefit-risk profile when compared to standard mitral valve surgery in a selected mitral valve patient population.

The EVEREST II RCT evaluated 279 patients in a prospective, randomized, active controlled, multi-center clinical trial to evaluate the safety and effectiveness of the MitraClip device in the treatment of moderate-to-severe (3+) or severe (4+) chronic mitral regurgitation (MR). Patients were randomized in a 2:1 fashion to receive either the MitraClip (Treatment) or mitral valve surgery (Control), respectively. The primary safety endpoint demonstrated a higher rate of adverse events for the surgical group. However, this composite of major adverse events (MAE) was driven by the need for transfusion (>2 units) in the surgery group. In addition, safety results in the surgical arm of the RCT were confounded by concomitant operations (including CABG, valve surgery, MAZE, PFO closure, ASD closure and LAA closure) in 47.5% of surgical patients. The sponsor statistically met the primary effectiveness endpoint using their definition of clinical success – freedom from death, MR>2+, and mitral valve re-intervention at 12 months. Nevertheless, the primary clinical effectiveness endpoint showed a large difference between the two treatments as the success rate was 88% in the surgery group and 72% in the percutaneous-repair group. Furthermore, the effectiveness results remain highly questionable for several reasons. FDA considers the prospectively defined primary effectiveness endpoint of clinical success as freedom from death, MR>1+, and mitral valve re-intervention at 12 months as a more appropriate endpoint, and notified the sponsor of this prior to the initiation of the trial. In the per protocol population, the MitraClip did not meet the FDA primary effectiveness endpoint (69% success for surgery and 45% for MitraClip). The effect on LV function as assessed by echocardiography also favored surgery by a wide margin. It should also be noted that 46% of patients in the MitraClip arm had moderate or greater MR (≥ 2+) whereas only 17% of surgical patients had MR ≥2+ at 12 months. By one year, 20% of MitraClip patients underwent MV surgery after MitraClip implantation, whereas only 2% of MV surgery patients underwent repeat MV

surgery. Finally, the heterogeneous patient population (e.g., MR etiology) studied in this moderately sized trial raises questions about the generalizability of the RCT results.

For a variety of reasons, the EVEREST II HRR single arm registry data are not easily interpretable.

The EVEREST II High Risk Registry (HRR) was a single arm registry which enrolled 78 patients as an adjunct dataset to the RCT trial. It was intended to provide a complementary dataset to a positive RCT trial. Since the EVEREST II RCT was a negative trial, evaluation of the HRR on its own is difficult due to multiple design limitations which resulted in the lack of an identifiable target population. Major problems include:

- Lack of an appropriate control or comparator group;
- Inclusion/exclusion criteria that did not adequately define a high risk patient population, much less an inoperable patient population as is described in the current indication for use requested for the MitraClip;
- Wide range in surgeon experience which resulted in increased variability in patient selection; and
- Heterogeneous patient population with regard to MR etiology.

Data presented for procedural mortality may appear to be better than the sponsor-defined surgeon predicted risk of mortality, but when one compares procedural mortality to objective STS scores for mitral valve replacement, the MitraClip does not show improvements in procedural mortality compared to that predicted for surgery by the STS score. It should be noted that relying on a risk prediction score as a comparator carries with it significant limitations, as these scores are much better served to help guide patient selection. If one were to conduct a comparison in procedural mortality using the STS score, however, this would be more appropriately done in relation to the STS score for mitral valve repair since this is a repair device being implanted in patients whose mitral valve anatomic and pathologic characteristics are overwhelmingly favorable for repair. The STS scores for mitral valve repair are consistently lower than the risk of mortality by at least 33% to 50% when compared to the STS score for mitral valve replacement. In addition, it should be noted that 18 of the 78 (23%) HRR patients died before 1 year and that 11 of those 18 (61%) deaths were CEC-adjudicated as possibly or probably related to the device or procedure. Differences in effectiveness measures (e.g., MR grade, LV function) were detected from baseline to 12 months, but are difficult to interpret without an appropriate comparator group in an unblinded single arm registry such as this.

REALISM HR is a continued access protocol cohort that was not intended to be used as a pivotal data set and is difficult to interpret.

The REALISM High Risk (HR) arm of the REALISM continued access protocol (CAP) study was also a single arm registry with the same inclusion/exclusion criteria as the HRR; therefore, the same concerns apply. It should be noted that interpretation of the REALISM

HR may be further complicated by the fact that many inoperable patients were treated as compassionate use patients concurrently with enrollment of the REALISM HR patients. This may have had the unintended consequence of skewing the REALISM HR patients towards a less sick population.

The Integrated High Surgical Risk Cohort, developed by pooling two registry data sets in a *post-hoc* manner, has major design limitations.

The Integrated High Surgical Risk Cohort was developed by pooling two cohorts, EVEREST II HRR and REALISM HR, in a *post-hoc* manner. The concerns raised above with respect to these two data sets are relevant to this analysis as well. It is difficult to conclude whether or not procedural mortality for the MitraClip was better than predicted surgical procedural mortality, since this depends greatly on whether subjective surgeon assessment of mortality, STS predicted rate of mortality for MV replacement, or STS score for MV repair (http://riskcalc.sts.org/STSWebRiskCalc273/) is used. Differences in effectiveness measures (e.g., MR grade, LV function, Quality of Life) were detected from baseline to 12 months, but are difficult to interpret without an appropriate comparator group in this cohort comprised of two pooled, unblinded, single arm registries. Interpretation of these data is further complicated by the heterogeneity of the patients enrolled (e.g., MR etiology, surgical risk). As expected with a *post-hoc* analysis, there are major design limitations when each individual data set has weaknesses to begin with. Thus, any inferences drawn from the Integrated High Surgical Risk Cohort to support the new *post-hoc* indication are highly questionable and should be more appropriately viewed as hypothesis generating.

The Duke Propensity Score Analysis was a retrospective, subset analysis with results that are difficult to interpret and where the matched cohorts do not represent any well-defined population.

The Duke Statistical Analysis Plans for the propensity score methodology (the original and its revisions) were not submitted for FDA review prior to the completion of the study. FDA has questions and concerns about the design and implementation of the propensity score methodology. It should be noted that the Duke Cohort of patients is a limited control group due to methodological reasons as well as limitations of the data contained in these analyses, which is further supported by the fact that no propensity score matching for endpoints other than mortality (e.g., MR grade, left ventricular function) was performed. Importantly, the propensity score matching did not provide a good match for all 211 MitraClip Propensity Score Analysis (PSA) Cohort patients. This is of concern to FDA as the make-up of the matched subjects (for Matched Cohort 1, n=127) does not represent any well-defined, prospectively selected population. We note that the Kaplan-Meier freedom from death data show an apparent benefit of the Matched Cohort 1 MitraClip patients over the Matched Cohort 1 Duke patients. However, when considering the MitraClip patients from Cohort 1 that were not matched, these patients have an STS score closer to the Duke patients and the Kaplan-Meier curves appear to overlap, indicating no mortality benefit. These concerns regarding the Duke Propensity Score Analyses make the data difficult to interpret and suggest that this analysis should be used mainly for hypothesis generation.

12 FDA RECOMMENDATION

The FDA believes that the analyses provided in the PMA are interesting and important. However, for the reasons discussed above, FDA believes these analyses are hypothesis generating and do not constitute valid scientific evidence of safety and effectiveness for the MitraClip CDS for the proposed Indication for use in an inoperable MR population.

Because there are limited options for high risk inoperable and high risk mitral valve patients, the FDA has recently worked with the sponsor in a highly interactive manner to develop a new U.S. randomized controlled trial for this patient population called the COAPT Trial. In addition, the sponsor has also recently developed a European randomized trial for this patient population.

FDA firmly believes that the currently enrolling COAPT and European trials are well-designed trials that will help to answer the many important questions posed by the very limited data analyses presented in this PMA and the mitral regurgitation literature. Therefore, FDA recommends that the MitraClip CDS continue to remain available to this vulnerable patient population as an investigational device so that Abbott Vascular can conduct the COAPT and European trials in an optimal manner. PMA approval is not appropriate at this time as major questions of safety and effectiveness, as well as the overall benefit-risk profile for this device, remain unanswered per the regulatory standards set forth in 21 CFR 860.7 (c)2, (d)1 and (e)1 (see Appendix A). Specifically, FDA cannot determine that in a significant portion of the target inoperable population, the use of the device for its intended uses and conditions of use, when accompanied by adequate directions for use and warnings against unsafe use, will provide clinically significant results without a valid comparator group, to support an indication for use for inoperable patients based on patients who were not necessarily inoperable.

13 Appendix A: 21CFR860.7

[Code of Federal Regulations]
[Title 21, Volume 8]
[Revised as of April 1, 2012]
[CITE: 21CFR860.7]

TITLE 21--FOOD AND DRUGS
CHAPTER I--FOOD AND DRUG ADMINISTRATION
DEPARTMENT OF HEALTH AND HUMAN SERVICES
SUBCHAPTER H--MEDICAL DEVICES

PART 860 -- MEDICAL DEVICE CLASSIFICATION PROCEDURES

Subpart A--General
Sec. 860.7 Determination of safety and effectiveness.

(a) The classification panels, in reviewing evidence concerning the safety and effectiveness of a device and in preparing advice to the Commissioner, and the Commissioner, in making determinations concerning the safety and effectiveness of a device, will apply the rules in this section.

(b) In determining the safety and effectiveness of a device for purposes of classification, establishment of performance standards for class II devices, and premarket approval of class III devices, the Commissioner and the classification panels will consider the following, among other relevant factors:

(1) The persons for whose use the device is represented or intended;

(2) The conditions of use for the device, including conditions of use prescribed, recommended, or suggested in the labeling or advertising of the device, and other intended conditions of use;

(3) The probable benefit to health from the use of the device weighed against any probable injury or illness from such use; and

(4) The reliability of the device.

(c)(1) Although the manufacturer may submit any form of evidence to the Food and Drug Administration in an attempt to substantiate the safety and effectiveness of a device, the agency relies upon only valid scientific evidence to determine whether there is reasonable assurance that the device is safe and effective. After considering the nature of the device and the rules in this section, the Commissioner will determine whether the evidence submitted or otherwise available to the Commissioner is valid scientific evidence for the purpose of determining the safety or effectiveness of a particular device and whether the available

evidence, when taken as a whole, is adequate to support a determination that there is reasonable assurance that the device is safe and effective for its conditions of use.

(2) Valid scientific evidence is evidence from well-controlled investigations, partially controlled studies, studies and objective trials without matched controls, well-documented case histories conducted by qualified experts, and reports of significant human experience with a marketed device, from which it can fairly and responsibly be concluded by qualified experts that there is reasonable assurance of the safety and effectiveness of a device under its conditions of use. The evidence required may vary according to the characteristics of the device, its conditions of use, the existence and adequacy of warnings and other restrictions, and the extent of experience with its use. Isolated case reports, random experience, reports lacking sufficient details to permit scientific evaluation, and unsubstantiated opinions are not regarded as valid scientific evidence to show safety or effectiveness. Such information may be considered, however, in identifying a device the safety and effectiveness of which is questionable.

(d)(1) There is reasonable assurance that a device is safe when it can be determined, based upon valid scientific evidence, that the probable benefits to health from use of the device for its intended uses and conditions of use, when accompanied by adequate directions and warnings against unsafe use, outweigh any probable risks. The valid scientific evidence used to determine the safety of a device shall adequately demonstrate the absence of unreasonable risk of illness or injury associated with the use of the device for its intended uses and conditions of use.

(2) Among the types of evidence that may be required, when appropriate, to determine that there is reasonable assurance that a device is safe are investigations using laboratory animals, investigations involving human subjects, and nonclinical investigations including in vitro studies.

(e)(1) There is reasonable assurance that a device is effective when it can be determined, based upon valid scientific evidence, that in a significant portion of the target population, the use of the device for its intended uses and conditions of use, when accompanied by adequate directions for use and warnings against unsafe use, will provide clinically significant results.

(2) The valid scientific evidence used to determine the effectiveness of a device shall consist principally of well-controlled investigations, as defined in paragraph (f) of this section, unless the Commissioner authorizes reliance upon other valid scientific evidence which the Commissioner has determined is sufficient evidence from which to determine the effectiveness of a device, even in the absence of well-controlled investigations. The Commissioner may make such a determination where the requirement of well-controlled investigations in paragraph (f) of this section is not reasonably applicable to the device.

(f) The following principles have been developed over a period of years and are recognized by the scientific community as the essentials of a well-controlled clinical investigation. They provide the basis for the Commissioner's determination whether there is reasonable assurance that a device is effective based upon well-controlled investigations and are also useful in

assessing the weight to be given to other valid scientific evidence permitted under this section.

(1) The plan or protocol for the study and the report of the results of a well-controlled investigation shall include the following:

(i) A clear statement of the objectives of the study;

(ii) A method of selection of the subjects that:

(a) Provides adequate assurance that the subjects are suitable for the purposes of the study, provides diagnostic criteria of the condition to be treated or diagnosed, provides confirmatory laboratory tests where appropriate and, in the case of a device to prevent a disease or condition, provides evidence of susceptibility and exposure to the condition against which prophylaxis is desired;

(b) Assigns the subjects to test groups, if used, in such a way as to minimize any possible bias;

(c) Assures comparability between test groups and any control groups of pertinent variables such as sex, severity or duration of the disease, and use of therapy other than the test device;

(iii) An explanation of the methods of observation and recording of results utilized, including the variables measured, quantitation, assessment of any subject's response, and steps taken to minimize any possible bias of subjects and observers;

(iv) A comparison of the results of treatment or diagnosis with a control in such a fashion as to permit quantitative evaluation. The precise nature of the control must be specified and an explanation provided of the methods employed to minimize any possible bias of the observers and analysts of the data. Level and methods of "blinding," if appropriate and used, are to be documented. Generally, four types of comparisons are recognized:

(a)No treatments. Where objective measurements of effectiveness are available and placebo effect is negligible, comparison of the objective results in comparable groups of treated and untreated patients;

(b)Placebo control. Where there may be a placebo effect with the use of a device, comparison of the results of use of the device with an ineffective device used under conditions designed to resemble the conditions of use under investigation as far as possible;

(c)Active treatment control. Where an effective regimen of therapy may be used for comparison, e.g., the condition being treated is such that the use of a placebo or the withholding of treatment would be inappropriate or contrary to the interest of the patient;

(d)Historical control. In certain circumstances, such as those involving diseases with high and predictable mortality or signs and symptoms of predictable duration or severity, or in the case of prophylaxis where morbidity is predictable, the results of use of the device may be

compared quantitatively with prior experience historically derived from the adequately documented natural history of the disease or condition in comparable patients or populations who received no treatment or who followed an established effective regimen (therapeutic, diagnostic, prophylactic).

(v) A summary of the methods of analysis and an evaluation of the data derived from the study, including any appropriate statistical methods utilized.

(2) To insure the reliability of the results of an investigation, a well-controlled investigation shall involve the use of a test device that is standardized in its composition or design and performance.

(g)(1) It is the responsibility of each manufacturer and importer of a device to assure that adequate, valid scientific evidence exists, and to furnish such evidence to the Food and Drug Administration to provide reasonable assurance that the device is safe and effective for its intended uses and conditions of use. The failure of a manufacturer or importer of a device to present to the Food and Drug Administration adequate, valid scientific evidence showing that there is reasonable assurance of the safety and effectiveness of the device, if regulated by general controls alone, or by general controls and performance standards, may support a determination that the device be classified into class III.

(2) The Commissioner may require that a manufacturer, importer, or distributor make reports or provide other information bearing on the classification of a device and indicating whether there is reasonable assurance of the safety and effectiveness of the device or whether it is adulterated or misbranded under the act.

(3) A requirement for a report or other information under this paragraph will comply with section 519 of the act. Accordingly, the requirement will state the reason or purpose for such request; will describe the required report or information as clearly as possible; will not be imposed on a manufacturer, importer, or distributor of a classified device that has been exempted from such a requirement in accordance with 860.95; will prescribe the time for compliance with the requirement; and will prescribe the form and manner in which the report or information is to be provided.

(4) Required information that has been submitted previously to the Center for Devices and Radiological Health, the Center for Biologics Evaluation and Research, or the Center for Drug Evaluation and Research, as applicable, need not be resubmitted, but may be incorporated by reference.

[43 FR 32993, July 28, 1978, as amended at 53 FR 11253, Apr. 6, 1988; 73 FR 49942, Aug. 25, 2008]