SUMMARY OF SAFETY AND EFFECTIVENESS DATA (SSED)

I. GENERAL INFORMATION

Device Generic Name: Dako anti-her2 IHC system

Device Trade Name: HercepTest™

Applicant's Name and Address: Dako Denmark A/S
Produktionsvej 42
DK-2600 Glostrup
Denmark

Date(s) of Panel Recommendation: None

Premarket Approval Application (PMA) Number: P980018/S010

Date of FDA Notice of Approval: October 20, 2010

Expedited: Not applicable.

The original PMA (P980018) for HercepTest™ was approved on 09/25/1998. This device is a semi-quantitative immunohistochemical assay to determine HER2 overexpression in breast cancer tissues routinely processed for histological evaluation. HercepTest is indicated as an aid in the assessment of patients for whom Herceptin® (transtuzumab) treatment is being considered.

The SSED to support the indication is available on the CDRH website and is incorporated by reference here. The current supplement was submitted to expand the indication for the HercepTest™ to metastatic gastric cancer patients.

II. INDICATIONS FOR USE

For in vitro diagnostic use.
HercepTest™ is a semi-quantitative immunocytochemical assay to determine HER2 protein overexpression in breast cancer tissues routinely processed for histological evaluation and formalin-fixed, paraffin-embedded cancer tissue from patients with metastatic gastric or gastroesophageal junction adenocarcinoma.

HercepTest™ is indicated as an aid in the assessment of patients for whom Herceptin® (trastuzumab) treatment is being considered (see Herceptin® package insert).

III. CONTRAINDICATIONS

None.

IV. WARNINGS AND PRECAUTIONS

The warnings and precautions can be found in the HercepTest™ labeling.

V. DEVICE DESCRIPTION

The device, HercepTest™ does not differ from the previously approved device as described in the original device description. The fundamental differences between the kit's use for breast and gastric cancer are procedural and recommendations with regard to specimen preparation, specimen evaluation, and result interpretation. These changes and mitigations are indicated below.

Specimen Preparation – Gastric
Adenocarcinoma specimens of the stomach, including gastroesophageal junction from biopsies, excisions, or resections must be handled correctly to preserve the tissue for immunohistochemical staining. Standard methods of tissue processing should be used for all specimens. When testing small biopsy specimens, ascertain intact tumor morphology and the presence of sufficient tumor cells for IHC evaluation. If HercepTest™ analysis is performed on a biopsy specimen, multiple (7-8) evaluable biopsy specimens from different regions of the tumor should be analyzed to ensure reliable determination of Her-2 status.

Interpretation of Staining - Gastric
Only specimens from patients with adenocarcinoma of the stomach, including gastroesophageal junction, should be scored. In cases with intestinal metaplasia and gastric adenocarcinoma in the same specimen, only the gastric adenocarcinoma component should be scored. For interpretation of HercepTest™ stained biopsies a cluster of at least 5 stained tumor cells is recommended.

Table 1. Interpretation and scoring of HER2 immunohistochemical staining

Score	Surgical Specimen – Staining Pattern	Biopsy Specimen – Staining Pattern	HER2 Overexpression Assessment
0	No reactivity or membranous reactivity in < 10% of tumor cells	No reactivity or no membranous reactivity in any (or < 5 clustered) tumor cell	Negative
1+	Faint/barely perceptible membranous reactivity in ≥ 10% of tumor cells; cells are reactive only in part of their membrane	Tumor cell cluster (≥ 5 cells) with a faint/barely perceptible membranous reactivity irrespective of percentage of tumor cells stained	Negative
2+	Weak to moderate complete, basolateral or lateral membranous reactivity in ≥ 10% of tumor cells	Tumor cell cluster (≥ 5 cells) with a weak to moderate complete, basolateral or lateral membranous reactivity irrespective of percentage of tumor cells stained	Equivocal

Score	Surgical Specimen – Staining Pattern	Biopsy Specimen – Staining Pattern	HER2 Overexpression Assessment
3+	Strong complete, basolateral or lateral membranous reactivity in ≥ 10% of tumor cells	Tumor cell cluster (≥ 5 cells) with a strong complete, basolateral or lateral membranous reactivity irrespective of percentage of tumor cells stained	Positive

Guidelines based on Hofmann et al. (1).
HercepTest™ is interpreted as negative for Her2 protein- overexpression (0 and 1+ score), equivocal (2+ score), and positive (3+ score).

Additional Recommendations for Interpretation of HercepTest™ Staining
Adenocarcinoma of the stomach, including gastroesophageal junction tested for HER2 protein overexpression are scored from 0 to 3+. While the 0 and 3+ cases are clear-cut, a small percentage of the remaining 1+ and 2+ samples may be more difficult to interpret. Use the following guidelines for interpretation of HercepTest™ staining in your laboratory.

- Evaluate the Control Cell Lines to validate the assay performance.
- Evaluate the Positive and Negative Control Slides.
- A hematoxylin and eosin (H&E) staining of the tissue specimen is recommended for the first evaluation. (The tumor may not be obvious when looking at the sample stained with HercepTest™. An H&E stained slide is required from the pathologist to verify the presence of the tumor). The HercepTestTM should be performed on a paired section (serial section) from the same paraffin block of the specimen.
- Evaluate the sections stained for HER2 protein overexpression at low power first. The majority of positive cases will be obvious at low power magnification.
- For 1+ cases, use 40x objective magnification to verify membrane staining.
- For 2+ cases, use 10x-20x objective magnification to verify membrane staining.

Surgical specimen
- Well-preserved and well-stained areas of the specimen should be used to make a determination of the percent of positive stained tumor cells.
- If a majority of tumor cells demonstrate complete, basolateral or lateral membrane staining, the staining is either 2+ or 3+.
- If there is complete, basolateral or lateral membrane staining at a strong intensity in equal to or more than 10% of the tumor cells in surgical specimens, the score of the specimen is 3+.
- If there is complete, basolateral or lateral membrane staining at a weak to moderate intensity in equal to or more than 10% of the tumor cells in surgical specimens, the score of the specimen is 2+.
- If equal to or more than 10% of the tumor cells in surgical specimens, stained only in part of their membrane, have a faint/barely perceptible intensity, the score of the specimen is 1+.
- If no staining is observed the score of the surgical specimen is 0.

- If less than 10% of the tumor cells in surgical specimens have staining, irrespective of the staining pattern (e.g. complete, basolateral, lateral or part of their membrane), the score is 0.

Biopsy specimen
- If there is a tumor cell cluster of at least 5 stained tumor cells with a strong complete, basolateral or lateral membrane staining, the score of the biopsy specimen is 3+, irrespective of percentage of tumor cells stained.
- If there is a tumor cell cluster of at least 5 stained tumor cells with a weak to moderate complete, basolateral or lateral membrane staining, the score of the biopsy specimen is 2+, irrespective of percentage of tumor cells stained.
- If there is a tumor cell cluster of at least 5 stained tumor cells with a faint/barely perceptible membrane staining and cells are stained only in part of their membrane, the score of the biopsy specimen is 1+, irrespective of percentage of tumor cells stained.
- If no staining is observed the score of the biopsy specimen is 0.
- If membrane staining (irrespective of staining intensity) is observed in less than 5 clustered tumor cells, the score of the biopsy specimen is 0.

VI.   ALTERNATIVE PRACTICES AND PROCEDURES

At present the recommended practice for HER2 testing includes immunohistochemical (IHC) staining for HER2 overexpression and in situ hybridization (ISH) testing for determination of gene copy number.

VII.   MARKETING HISTORY

HercepTest™ for the extended indication in gastric cancer has been marketed in the European Union countries, since March 2010 and Canada since April 2010. The product has not been withdrawn from marketing for any reason related to its safety or effectiveness.

VIII.   POTENTIAL ADVERSE EFFECTS OF THE DEVICE ON HEALTH

HercepTest™ is intended for in vitro diagnostic use only. As a consequence, there is no potential direct adverse effect on the patient's health. Any potential adverse effects would be related to misuse of the device or interpretation error leading to potentially incorrect diagnosis and therapy selection.

IX.   SUMMARY OF PRECLINICAL STUDIES

A. Laboratory Studies
1. Non-Clinical Studies - Internal
a. Analytical Sensitivity on Gastric Cancer Specimens
The analytical sensitivity of HercepTest™ (HER2 antibody) when used on gastric cancer tissue was investigated using 22 different gastric

adenocarcinoma specimens tested on three lots of HercepTest™ (3 * 22 specimens). The result of the analytical sensitivity demonstrated HercepTest's ability to consistently detect the target substance (HER2 protein) in gastric adenocarcinoma specimens with various level of HER2 expression (negative (0, 1+), equivocal (2+), and positive (3+).

b. **Analytical Specificity on Gastric Cancer Specimens**
Analytical specificity was tested using 22 different gastric adenocarcinoma specimens (one specimen/slide) using three lots of HercepTest™. HercepTest™ (HER2 antibody) specifically detected the HER2 protein localized in the cell membrane. The observed HER2 staining pattern on gastric adenocarcinoma tissue specimens were in accordance to the staining pattern described in Hoffmann et al, 2008.[1]

c. **Robustness on gastric cancer specimens**
The robustness of HercepTest™ on gastric adenocarcinoma specimens was tested in the epitope retrieval step as the other steps in the HercepTest™ staining procedure is strictly defined. The epitope retrieval step in the HercepTest™ staining procedure, allows incubation for 40 (± 1) minutes at 95-99 °C in Epitope Retrieval Solution, followed by cooling of the slides in the Epitope Retrieval Solution for 20 (± 1) minutes at room temperature. After epitope retrieval, slides are soaked in Wash Buffer for 5-20 minutes prior to staining.

The robustness of the epitope retrieval step was tested by varying epitope retrieval times and temperatures, prior to the 20 minute cooling step in the Epitope Retrieval Solution, and varying the washing times afterwards, in the staining procedure.

No difference in results was observed at the following experimental conditions:
- Target retrieval - 40 min. at 99 °C and soak in Wash Buffer for 5 min.
- Target retrieval - 40 min. at 99 °C and soak in Wash Buffer for 20 min.
- Target retrieval - 40 min. at 97 °C and soak in Wash Buffer for 5 min.
- Target retrieval - 40 min. at 97 °C and soak in Wash Buffer for 20 min.
- Target retrieval - 40 min. at 95 °C and soak in Wash Buffer for 5 min.
- Target retrieval - 40 min. at 95 °C and soak in Wash Buffer for 20 min.

d. **Repeatability on gastric cancer specimens**
The repeatability of the HER2 score was investigated with HercepTest™ using three consecutive sections from 11 different gastric adenocarcinoma specimens. The repeatability study met the acceptance criteria and demonstrated agreement between repeated assessments of HER2 score carried out under the same conditions.

2. Non-Clinical Studies – External

a. Reproducibility Study

The day-to-day, site-to-site, observer-to-observer and automated-to-manual reproducibility study of HercepTest™ was performed on gastric adenocarcinoma specimens from the stomach or gastroesophageal junction (GEJ).

Study Design

The study involved three sites using blinded surgical resections and biopsy substitutes of formalin-fixed, paraffin-embedded (FFPE) human adenocarcinoma tissue specimens (60) from stomach or GEJ. At each study site sections from 60 different specimens were stained and analyzed on five non-consecutive days. The specimens represented 20 HER2 negative (IHC 0 or IHC 1+), 20 HER2 equivocal (IHC 2+) and 20 HER2 positive (IHC 3+) cases. The total number stained sections are shown in Table 2. HER2 status was defined as negative (HercepTest™ scores of 0 or 1+), as equivocal (HercepTest™ score 2+) and as positive (HercepTest™ score of 3+).

Table 2. Number of sections including control sections stained

	Site 1		Site 2		Site 3		Total
	HER2	NegC	HER2	NegC	HER2	NegC	
Run #1	60	60	60	60	60	60	360
Run #2	60	0	60	0	60	0	180
Run #3	60	0	60	0	60	0	180
Run #4	60	0	60	0	60	0	180
Run #5	60	0	60	0	60	0	180
Run #6M	60	60					120
Run #8M	60	0					60
Total	420	120	300	60	300	60	1260

In Run #6 and Run #8 sections are stained manually, remaining runs are automated.
NegC = Negative Control Reagent.

Results

The day-to-day variation within the three study sites was determined by calculating HER2 status agreement from cross tabulations for every possible comparison of days within the sites (see Table 3).

Table 3. Day-to-Day Agreements.

		Overall Agreement (%)			
		Observer 1		Observer 2	
		Agreement	95% CI Lower Limit	Agreement	95% CI Lower Limit
Site 1	Day 1 vs Day 2	85.0	74.4	93.3	84.9
	Day 1 vs Day 3	90.0	80.5	96.7	89.7
	Day 1 vs Day 4	83.3	72.4	93.3	84.9

	Day 1 vs Day 5	85.0	74.4	88.3	78.5
	Day 2 vs Day 3	95.0	87.3	96.7	89.7
	Day 2 vs Day 4	91.7	82.7	93.3	84.9
	Day 2 vs Day 5	90.0	80.5	91.7	82.7
	Day 3 vs Day 4	93.3	84.9	93.3	84.9
	Day 3 vs Day 5	95.0	87.3	91.7	82.7
	Day 4 vs Day 5	91.7	82.7	91.7	82.7
Site 2	Day 1 vs Day 2	95.0	87.3	83.1	72.0
	Day 1 vs Day 3	96.7	89.7	84.7	74.0
	Day 1 vs Day 4	96.7	89.7	86.4	76.0
	Day 1 vs Day 5	91.7	82.7	83.1	72.0
	Day 2 vs Day 3	98.3	92.5	95.0	87.3
	Day 2 vs Day 4	98.3	92.5	90.0	80.5
	Day 2 vs Day 5	96.7	89.7	90.0	80.5
	Day 3 vs Day 4	96.7	89.7	95.0	87.3
	Day 3 vs Day 5	95.0	87.3	95.0	87.3
	Day 4 vs Day 5	95.0	87.3	93.3	84.9
Site 3	Day 1 vs Day 2	90.0	80.5	91.7	82.7
	Day 1 vs Day 3	91.7	82.7	91.7	82.7
	Day 1 vs Day 4	88.3	78.5	96.7	89.7
	Day 1 vs Day 5	88.3	78.5	85.0	74.4
	Day 2 vs Day 3	95.0	87.3	90.0	80.5
	Day 2 vs Day 4	91.7	82.7	95.0	87.3
	Day 2 vs Day 5	91.7	82.7	90.0	80.5
	Day 3 vs Day 4	96.7	89.7	91.7	82.7
	Day 3 vs Day 5	93.3	84.9	90.0	80.5
	Day 4 vs Day 5	90.0	80.5	88.3	78.5

The vast majority of comparisons showed an overall agreement at 90.0% or above, and in 20 out of the 60 comparisons the overall agreement was found at 95.0% or above. Using Fisher's exact test it was found that results were reproducible between days for each specimen category. A comparison for all of the sections of a specimen over all runs was calculated using the Gini index, to determine the probability of disagreement between two categorical results (Agresti, *Categorical Data* Analysis, 2002; Bishop, Feinberg, Holland, *Discrete Multivariate Analysis*, 1975).

An analysis of the overall variability by specimen category (negative, equivocal and positive) was provided separately for each of the 20 specimen within the categories and is summarized below.

Category=Negative

Tissue	location	range of Gini Index	Median
Biopsy	GEJ	0.00-0.00	0.00
Biopsy	Stomach	0.00-0.42	0.00
Surgical	GEJ	0.00-0.28	0.00

Surgical   Stomach   0.00-0.06   0.00
Average Gini Index for this category =0.04
Median Gini Index for this category=0.00

Category=Equivocal

Tissue	location	range of Gini Index	Median
Biopsy	GEJ	0.06-0.46	0.21
Biopsy	Stomach	0.00-0.49	0.44
Surgical	GEJ	0.34-0.52	0.43
Surgical	Stomach	0.06-0.48	0.18

Average Gini Index for this category =0.29
Median Gini Index for this category=0.35

Category=Positive

Tissue	location	range of Gini Index	Median
Biopsy	GEJ	0.00-0.38	0.15
Biopsy	Stomach	0.00-0.46	0.25
Surgical	GEJ	0.00-0.46	0.03
Surgical	Stomach	0.00-0.44	0.00

Average Gini Index for this category =0.13
Median Gini Index for this category=0.00

The value of Gini index ranges from 0 to 1 where a Gini Index of 0 indicates that the probability that readings from a tissue sample fall into different category is 0, indicating a good agreement and 1 indicates that the probability that readings from a tissue sample fall into different category is 1 indicating a poor agreement. Note that the median of the Gini Index for both the negative and positive category indicates a very good agreement in these two categories.

The site-to-site variation was determined for the first observer by calculating the overall HER2 status agreement from 3x3 cross tabulations. The overall average agreements observed were 82.7%, 75.0% and 88.0% between the three sites (see Table 4).

Table 4. Site-to-site agreements

		Overall Agreement (%)		
		Agreement	95% CI Lower Limit	Average Agreement
Site 1 vs. 2	Day 1 vs Day 1	83.3	72.4	82.7
	Day 2 vs Day 2	85.0	74.4	
	Day 3 vs Day 3	85.0	74.4	
	Day 4 vs Day 4	81.7	70.5	
	Day 5 vs Day 5	78.3	66.7	
Site 1 vs. 3	Day 1 vs Day 1	80.0	68.6	75.0
	Day 2 vs Day 2	73.3	61.2	
	Day 3 vs Day 3	78.3	66.7	
	Day 4 vs Day 4	68.3	55.9	

12

	Day 5 vs Day 5	75.0	63.0	
Site 2 vs. 3	Day 1 vs Day 1	88.3	78.5	
	Day 2 vs Day 2	86.7	76.4	
	Day 3 vs Day 3	90.0	80.5	88.0
	Day 4 vs Day 4	86.7	76.4	
	Day 5 vs Day 5	88.3	78.5	

Two manual runs were performed at site one to allow for comparison between HER2 status obtained by automated and manual staining platforms. Agreements were calculated from 3x3 cross tabulations and were at or above 90% (see Table 5).

Table 5. Automated-to-manual agreements

		Overall Agreement (%)		
		Agreement	95% CI Lower Limit	Average Agreement
Site 1, Observer 1	Day 1 vs Manual 6	86.7	76.4	90.0
	Day 2 vs Manual 8	93.3	84.9	
Site 2, Observer 2	Day 1 vs Manual 6	90.0	80.5	90.9
	Day 2 vs Manual 8	91.7	82.7	

Agreement between the two observers was determined for each run at the three sites. The average observer to observer agreement was 88.0% at site 1, 83.6% at site 2 and 81.0% at site 3 (see Table 6).

Table 6. Observer-to-Observer Agreements

Observer one vs Observer two		Overall Agreement (%)		
		Agreement	95% CI Lower Limit	Average Agreement
Site 1	Day 1	91.7	82.7	
	Day 2	91.7	82.7	
	Day 3	93.3	84.9	88.0
	Day 4	83.3	72.4	
	Day 5	80.0	68.6	
Site 2	Day 1	86.4	76.0	
	Day 2	83.3	72.4	
	Day 3	83.3	72.4	83.6
	Day 4	83.3	72.4	
	Day 5	81.7	70.5	
Site 3	Day 1	80.0	68.6	
	Day 2	78.3	66.7	
	Day 3	80.0	68.6	81.0
	Day 4	78.3	66.7	
	Day 5	90.0	80.5	

In conclusion the HercepTest™ analysis of 60 different gastric cancer specimens, obtained from stomach or gastroesophageal junction performed on five non-consecutive days at three study sites based on HER2 status showed overall day-to-day agreements in the range of 83.1% to 98.3% in the 60 comparisons. The vast majority of comparisons (47 of 60) resulted in overall agreements of 90.0% or above and Fisher's exact test revealed that the observed results were not different between days. Site-to-site agreement was in the range of 68.3% to 90.0%. The average overall agreements for the three possible site comparisons were 82.7%, 75.0% and 88.0%. According to Fisher's exact test, the results were not different between sites. Agreements between observers at each site were 88.0%, 83.6% and 81.0% for the three sites, and comparison of automated and manual HercepTest™ platforms revealed average overall agreements at or above 90.0%.

B. Animal Studies

None.

C. Additional Studies

1. Heterogeneity Analysis

To address the questions related to the heterogeneous nature of gastric cancer and the use of biopsy cores in the clinic for evaluation of HER2 status two additional assessments were performed:

- A heterogeneity assessment of the selected specimens from the clinical study BO18255 (ToGA trial) performed at central laboratory Targos Molecular Pathology Gmbh, D-34119 Kassel, Germany on biopsy specimens from the clinical trial.
- Study performed at Dako on the heterogeneity within a tissue block/section and the number of biopsy cores that should be analyzed in order to obtain a reliable result relative to the complete tumor.

An overview of the studies and assessment of the IHC specimens was provided and is summarized below.

a. Heterogeneity Study Performed at Targos Molecular Pathology Gmbh

A heterogeneity assessment study was performed at Targos central laboratory. The assessment was performed on selected IHC and FISH specimens from BO18255. Only results of the assessment of IHC specimens are covered in this submission. The evaluated IHC slides from BO18255 trial were selected based on the IHC HER2 score: IHC HER2 score 0/1+ (n=10), IHC HER2 score 2+ (n=20), and IHC HER2 score 3+ (n=10). Half of the cases were from stomach, while the other half of the cases were from GEJ.

For IHC, heterogeneity is defined as individual pieces having different HER2 score in a biopsy sample, i.e., if biopsy pieces have different scores then it was considered heterogeneous and if only one score for all pieces then it was not heterogeneous. The number of biopsies on the slide and number of evaluable biopsies on the slide were provided. The number of evaluable biopsy pieces in

each sample varies due to the followings: 1) not all biopsy pieces were present on the IHC slide or 2) no relevant tissue was available for some pieces of

Forty-five percent (9/20) of GEJ cases showed heterogeneity on the slide and were limited to IHC 1+ and IHC 2+ cases. Two of the GEJ cases were discarded because one contained only one biopsy and the second contained only one biopsy specimen that was evaluable however they were not discarded from the calculations. Thirty percent (6/20) of the stomach biopsy cases showed heterogeneity on the slide and included IHC 1+, 2+, and 3+ scores. The distribution of biopsies and IHC scores observed are summarized in Table 8.

Table 8. Distribution of biopsies and IHC scores observed

	Total specimens	Total bx	Eval bx	Heterogeneity observed (%	# specimens – categorization of IHC results observed					
					0/1+	0/2+	0/3+	0/1+/2+	1+/2+	1+/2+/3+
Stomach	20	87	72	6 (30%)	1	1	1	0	2	1
GEJ	20	115	90	9 (45%)	2	6	0	1	0	0

Thirty-seven and a half percent (15/40) of tumors in the stomach and GEJ were identified as heterogeneous. It is unclear however how large the biopsies for each specimen were. The biopsy cases in the BO18255 trial included 1-9 evaluable biopsy specimens. Heterogeneity on the slide in this study relates to the heterogeneity at the gross tumor level (i.e. sampling from different locations in the tumor) and both tumors from stomach and GEJ exhibited tumor heterogeneity (30% in stomach and 45% in GEJ). The conclusion is that multiple biopsy pieces should be evaluated for reliable HER2 status determination due to heterogeneity issue. A large number of the specimens are noted to demonstrate "no unequivocal strong membrane staining due to edge artifacts and crushing" so it is unclear how reliable the scores or assessment heterogeneity are.

b. **Heterogeneity Study Performed at Dako Denmark A/S - HercepTest™**
A second study titled "Evaluation of specimen size in gastric cancer" was performed at Dako Denmark. It was designed to determine the smallest amount of tissue from which a reliable result could be determined and related to the heterogenic nature of gastric cancer and the use of biopsy specimens. Seventy-five percent of the specimens included in the ToGA trial consisted of biopsy specimens.

Table 9 lists the specimens used in this study by primary site and HER2 status. The specimens represent both adenocarcinomas from stomach (13 specimens) and GE-junction (11 specimens) were from 24 different patients and include all four HER2 scores (0, 1+, 2+ and 3+). The 4 μm cut sections stained slides were categorized as Negative (HER2 score 0-1+), equivocal (HER2 score 2+) and positive (HER2 score 3+).

Table 9. Specimens included in Dako study.

HER2 Score	Stomach	GEJ	Total
0/1+	4	3	7
2+	5	5	10
3+	4	3	7
Total	13	11	24

The tumor area was divided into 2 mm x 2 mm squares in order to mimic biopsy samples and 9 squares were randomly selected were evaluated for each specimen. Unique random numbers within the number of squares possible were selected using the RAND function of Excel. In most specimens (21 out of 24) one or several squares representing the edge have been evaluated and no edge artifacts that influenced scoring were observed. The number of squares for all specimens ranged from 13-121.

The whole specimen was scored according to the scoring system for surgical specimens (10% cut-off for stained tumor cells) and squares were scored according to the scoring system for biopsy specimens (a cluster of at least 5 stained tumor cells). For determination of HER2 protein expression, only the membrane staining intensity and pattern was evaluated using the interpretation guidelines. Slide evaluation was performed by qualified trained personnel using a light microscope. Cytoplasmic staining was considered non-specific staining and was not included in the assessment of membrane staining intensity.
Tumor margins were included to mimic biopsies partially containing tumor. Squares were scored according to the guidelines established for biopsies as described in Hofmann et al. The total number tumor cells and the number of stained tumor cells per square (<6, 6-50, 51-100, 100-200, 200-500, >500) were recorded along with the staining intensity.

The evaluation of individual squares corresponds to evaluation of a biopsy-sample containing up to 9 biopsies. The study was designed to determine the fewest number of biopsies necessary to correspond to the full surgical section HER2 status. A combined score was evaluated for the highest HER2 score for 3, 6, and 9 squares. For example if the 3 squares have the following scores (2+, 2+, 3+) the combined score (reported score) is 3+. For each specimen it was determined whether the combined score of square (1-3), (1-6) and (1-9) correspond to the score of the surgical specimen and the percentage of specimens where the combined score matched the status for the whole slide was stated. In most, 87.5% (21/24) specimens one or several squares representing the margin were evaluated and no edge artifact that influenced scoring was observed. Dako initially determined that 84.6% (11/13) of the stomach biopsies and 81.8% (9/11) required a minimum of three (3) biopsies cores would be necessary.

16

Based on the data presented, Dako Denmark recommended that 3-8 biopsies be included in the determination of HER2 IHC status; however it was observed in several cases that a single change in squares assessed could change the combined score and their evaluation was based on sequential squares and not potential randomization of squares.

Repeat of Dako Heterogeneity study - HercepTest™
In order to confirm that the proposed labeling language "If HercepTest™ analysis is performed on a biopsy specimen, multiple (3-8) evaluable biopsy specimens from different regions of the tumor should be analyzed to ensure reliable determination of HER2 status" is appropriate. To minimize the time necessary to perform the study, it was agreed that Dako would perform a reassessment of heterogeneous specimens from the original study on the same heterogenous stained slides, however in 9 new randomly selected squares in the same manner as previously. To not bias the choice of squares, none of the previous squares were removed from consideration. As a result, in several cases the same squares were selected. Scoring, data collection and data evaluation were performed identically to the original study. The results showed heterogeneity in specimens from both stomach and GE junction and also that it is necessary to analyze several biopsy specimens from each patient case to reach a reliable result.

Results
One stomach specimen was originally classified as heterogeneous, however upon reassessment the specimen became non-heterogeneous. In the original study eight squares out of nine were score 0 and one square was scored 1+. In the reassessment study nine out of nine squares were scored 0. GEJ specimen that was originally classified as heterogeneous became non-heterogeneous upon re-evaluation. In the original study seven squares out of nine were score 3+ and two squares were scored 2+. In the reassessment study nine out of nine squares were scored 3+.

Based on the results of the reassessment study 14 out of 17 specimens were classified appropriately after 1-3 squares and three specimens were not classified correctly until squares 4-5 were scored. The reassessment of the original study showed that 15 out of 17 specimens (selected based on the original study) exhibited heterogeneity in HER2 status between squares and/or the surgical specimen. This indicates that in gastric cancer both stomach and GEJ specimens demonstrate heterogeneity in HER2 expression pattern and that several biopsies should be evaluated for each patient case to obtain a reliable HER2 result.

Due to different cut-offs for surgical specimens (10% of stained tumor cells) and biopsy specimens (a cluster of at least 5 stained tumor cells) it was observed that two surgical specimens that were categorized as equivocal were categorized as positive when evaluating squares (biopsy pieces). The study at

Targos showed that 15 out of 40 patient cases exhibited "within tumor heterogeneity" (both stomach and GEJ) when using 1-9 evaluable biopsy pieces in each patient case.

Conclusion – Heterogeneity studies
Based on the results of the original study, the originally proposed 3-8 biopsy specimens were recommended was revised to 6-8 evaluable biopsies from different regions of the tumor should be analyzed to ensure reliable determination of HER2 status. This range was still of concern and to err on the side of caution, it was requested that the recommendation to the user be increased slightly to a recommendation of 7-8 biopsies.

X.   SUMMARY OF PRIMARY CLINICAL STUDY

The BO18255 clinical trial (ToGA) established a reasonable assurance of safety and effectiveness with regards to HER2 testing when using *HER2* FISH pharmDx™ Kit for the assessment of patients with adenocarcinoma of the stomach, including gastro-esophageal junction, for whom trastuzumab (Herceptin®, Roche) is being considered. The study was conducted by F. Hoffmann-La Roche AG in the period of September 2005 to January 2009, with one year of additional follow-up to collect safety information through January 2010 and submitted to CDER under supplemental BLA 103792/5250 Herceptin in Gastric Adenocarcinoma.

A. Study Design
The BO18255 study *"An open-label randomized multicenter phase III study of trastuzumab in combination with a fluoropyrimidine and cisplatin versus chemotherapy alone as first-line therapy in patients with HER2 positive advanced gastric cancer"* was designed as a prospective, randomized, open-label, multi-center, Phase III study evaluating the efficacy of trastuzumab in combination with chemotherapy versus chemotherapy alone. After having fulfilled the protocol-defined screening for eligibility, including confirmation of HER2 positive status, the patients were randomized to treatment with trastuzumab plus fluoropyrimidine/cisplatin (FC+H), or fluoropyrimidine/cisplatin (FC) treatment arm in a 1:1 ratio. HER2 status was assessed by both fluorescence in situ hybridization, FISH, (*HER2* FISH pharmDx™ Kit, Dako) and by immunohistochemistry, IHC, (HercepTest™, Dako), and study eligibility required tumors to be either FISH+ or IHC3+.

Study start was September 2005 and the clinical data cutoff date for the definitive analysis of study outcomes was January 7, 2009. Patient enrollment was completed in December 2008. The database for this PMA supplement reflected data collected through January 7, 2009 and included 594 patients. The study was a non-U.S. study conducted in 24 countries, which included the following parts of the world: Asia, Australia, Europe, South and Central America, Russia, and South Africa. The HER2 analyses of the tumor specimens were performed at one single central laboratory (Targos Molecular Pathology GmbH, D-34119 Kassel, Germany). Treatment randomization in the study was stratified by Eastern Cooperative Oncology Group

18

(ECOG) performance status (PS), chemotherapy regimen (capecitabine versus 5-fluorouracil), locally advanced versus metastatic disease, primary origin in stomach versus gastro-esophageal junction, and measurable versus non-measurable disease.

The main efficacy outcome measure of the study was duration of overall survival (OS), defined as the time from the date of randomization to the date of the death (from any cause). For time to event endpoint, comparisons were made between treatment arms using the two-sided unstratified log-rank test. Kaplan-Meier curves, median and 95% confidence intervals (CI) were provided for each treatment arm as well as hazard ratio and its two-sided 95% CI from Cox regression were provided. Stratified analyses were also performed.

1. **Clinical Inclusion and Exclusion Criteria**
   Enrollment in the BO18255 study was limited to patients who met the following inclusion criteria: Patients with histologically confirmed inoperable locally advanced, recurrent and/or metastatic adenocarcinoma of the stomach including gastroesophageal junction, who had not been previously treated for their advanced/metastatic disease and whose tumors were HER2 positive either by IHC (3+) or FISH (*HER2*/CEN-17 ratio ≥ 2.0), were eligible for enrollment in the study. The HER2 status was assessed in a central laboratory by two methods in parallel, IHC and FISH. The tissue used for testing was either surgical resection or biopsies specimens.

   Patients were not permitted to enroll in the BO18255 study if they met any of the following exclusion criteria: Patients with previous chemotherapy for advanced/metastatic disease (prior adjuvant/neoadjuvant therapy was allowed, if at least 6 months had elapsed between completion of adjuvant/neoadjuvant therapy and enrollment; adjuvant/neoadjuvant therapy with a platin was not allowed), patients with active (significant or uncontrolled) gastrointestinal bleeding, patients with other malignancy within the last 5 years, except for carcinoma in situ of the cervix, or basal cell carcinoma.

2. **Follow-up Schedule**
   After enrollment, patients were to be administered 6 cycles of cytotoxic chemotherapy in both treatment arms, unless disease progression or intolerable toxicity occurred sooner. Patients in the experimental arm continued to be treated with trastuzumab after the completion of cytotoxic chemotherapy, until disease progression. Patients in both arms were assessed until disease progression, unacceptable toxicity or consent withdrawal. After progression, they were monitored for survival at regular 6 week intervals, until death or the study end (which was January 7, 2010).

3. **Clinical Endpoints**
   The main efficacy outcome measure of the study was OS, defined as the time from the date of randomization to the date of the death (from any cause).

19

B. Accountability of PMA Cohort

A total of 594 patients were enrolled to the study (Intent to Treat, ITT), 296 patients were randomized to the FC arm and 298 patients to the FC+H arm. A total of 10 randomized patients (N=6 FC and N=4 FC+H) did not receive any study drug and were determined to be non-eligible or declined to participate in the study after randomization but before treatment began.

C. Study Population Demographics and Baseline Parameters

The BO18255 study was conducted outside the USA at 122 sites in 24 countries in Asia, Australia, Europe, South and Central America, Russia, and South Africa. Based on the below presentation of the demographic data, it is seen that the Study BO18255 population is largely comparable to the U.S. population with advanced gastric cancer in terms of patient age, primary tumor site, extent of tumor, and type of cancer (adenocarcinoma). The clinical benefit observed in Study BO18255 was generally consistent across demographic subgroups in the study. Characteristics of the total U.S. general population as defined by the U.S. Census Bureau 2006 estimate (U.S. Total), the U.S. population with advanced gastric cancer, and the ITT population with advanced gastric cancer from the BO18255 study are shown in Table 10.

Table 10. Characteristics of Populations with Advanced Gastric Cancer

Demographic Characteristic	US Total Population[a] (n=299.4M)	US Advanced Gastric Cancer Population[b] (n=6,395)	Study BO18255 Advanced Gastric Cancer[c] (n=594)
Race			
White/Caucasian	73.9%	73.8%	37.7%
Black/African-American	12.4%	11.5%	0.5%
Asian	4.4%	14.3%	52.9%
Other (incl. multiracial)	9.3%	1.4%	8.9%
Sex			
Female	49.2%	62.1%	76.3%
Male	50.8%	37.9%	23.7%
Age			
Median	36.4 yrs	67.0 yrs	60.0 yrs
Mean	–	65.3 yrs	59.0 yrs
Primary site			
GE junction	–	73.7%	81.6%
Stomach	–	26.3%	18.4%
Extent of disease			
Locally advanced	–	5.0%	3.4%
Metastatic	–	95.0%	96.6%
Histology			
Adenocarcinoma	–	97.6%	100.0%
Other	–	2.4%	0.0%

Compared with the total U.S. general population, there is a higher proportion of Asians among U.S. patients with advanced gastric cancer (14.3% vs. 4.4%), reflecting a 3.25-fold increased risk and the higher incidence of advanced gastric cancer in the U.S. Asian population. Asian subjects were over-represented in the BO18255 study population compared with the U.S. population with advanced gastric cancer (52.9% vs. 14.3%). The greater proportion of males in the U.S. population with advanced gastric cancer, compared with the total U.S. population (62% vs. 49%) and older median age (67 vs. 36 years) indicate both sex and older age as possible risk factors for advanced gastric cancer. There was a larger fraction of male patients in the BO18255 study population compared with the U.S. population with advanced gastric cancer (76% vs. 62%).

The patient demographics of the study population are shown in Table 11. These characteristics were well-balanced across the two treatment arms. The study population comprised more males than females (76% vs. 24%). The majority of the population was oriental (54% in FC arm, 52% in FC+H arm) and the median age was 59 years in the FC arm and 61 years in the FC+H arm.

Table 11. Summary of Patient Demographic Data

	FC Arm (N=296)	FC+H Arm (N=298)	Total
Age (yr)			
Mean (SD)	58.5 (11.1)	59.4 (10.8)	59.0 (10.9)
Median	59	61	60
Range	21 - 82	23 - 83	21 - 83
Sex			
Female	73 (24.7%)	68 (22.8%)	141 (23.7%)
Male	223 (75.3%)	230 (77.2%)	453 (76.3%)
World region			
Asia	166 (56.1%)	158 (53.0%)	324 (54.5%)
C/S America	26 (8.8%)	27 (9.1%)	53 (8.9%)
Europe	95 (32.1%)	99 (33.2%)	194 (32.7%)
Other	9 (3.0%)	14 (4.7%)	23 (3.9%)
Race			
Black	2 (0.7%)	1 (0.3%)	3 (0.5%)
Caucasian	109 (36.8%)	115 (38.6%)	224 (37.7%)
Asian	160 (54.1%)	154 (51.7%)	314 (52.9%)
Other	25 (8.4%)	28(9.4%)	53 (8.9%)

2(

The stratification factors were well-balanced between the treatment arms as shown in Table 12. Overall, there were a high percentage of patients with metastatic disease (97%) and the primary site was mainly the stomach (82%). The majority of patients had an ECOG performance status of 0-1 (90%). For the majority of patients (87%), the chemotherapy regimen included capecitabine rather than 5-FU.

The baseline disease characteristics are summarized in Table 13. The median time from first diagnosis of gastric cancer to randomization was 1.2 months for the FC arm and 1.5 months for the FC+H arm. Less than 1% (4/594) of patients had prior anthracycline therapy, 2% (12/594) had prior radiotherapy, and 23% (135/594) had prior gastrectomy.

Table 12. Summary of Stratification Factors

	FC Arm (N=296)	FC+H Arm (N=298)
ECOG performance status		
0-1	269 (90.9%)	268 (89.9%)
2	27 (9.1%)	30 (10.1%)
Extent of disease		
Locally advanced	10 (3.4%)	10 (3.4%)
Metastatic	286 (96.6%)	288 (96.6%)
Primary site		
GE junction	51 (17.2%)	58 (19.5%)
Stomach	245 (82.8%)	240 (80.5%)
Measurability		
Measurable disease	263 (88.9%)	272 (91.3%)
Non-measurable disease	33 (11.1%)	26 (8.7%)
Chemotherapy regimen		
5-FU	36 (12.2%)	38 (12.8%)
Capecitabine	260 (87.8%)	259 (87.2%)

Table 13. Summary of Baseline Disease Characteristics

Characteristic	FC (n=296)	FC+H (n=298)
Time from first diagnosis of gastric cancer to randomization (mo)		
Mean (SD)	4.0 (8.3)	7.3 (21.4)
Median	1.2	1.5
Range	0 - 66	0 - 309
Time from diagnosis of locally advanced or recurrent/metastatic disease to randomization (mo)		
Mean (SD)	1.1 (0.8)	1.6 (2.4)

Median	1.0	1.0
Range	0 - 7	0 - 27
Type of gastric cancer (by central laboratory assessment)		
Diffuse	25 (8.4%)	26 (8.7%)
Intestinal ·	218 (73.6%)	227 (76.2%)
Mixed	50 (16.9%)	44 (14.8%)
Not assessed	3 (1.0%)	1 (0.3%)
Visceral (lung or liver) metastasis		
Yes	175 (59.1%)	170 (57.0%)
No	121 (40.9%)	128 (43.0%)
Prior gastrectomy		
Yes	63 (21.3%)	72 (24.2%)
No	233 (78.7%)	266 (75.8%)
Prior chemotherapy		
Yes	13 (4.4%)	27 (9.1%)
No	283 (95.6%)	271 (90.9%)
Number of metastatic sites		
N	295	296
1-2	149 (50.5%)	153 (51.7%)
>2	146 (49.5%)	143 (48.3)
Number of Metastatic lesions		
N	295	296
1-4	119 (40.3%)	129 (43.6%)
>4	176 (59.7%)	167 (56.4%)

n= for each group is considered to be 296 and 298, respectively unless otherwise specified

D. Safety and Effectiveness Results

The safety with respect to treatment with FC and the FC+H arms will not be addressed in the SSED for *HER2* FISH pharmDx™ Kit.

The main outcome measure of Study 7 was overall survival (OS), analyzed by the unstratified log-rank test. The final OS analysis based on 351 deaths was statistically significant (nominal significance level of 0.0193). An updated OS analysis was conducted at one year after the final analysis. The efficacy results of both the final and the updated analyses are summarized in Figure 1 and Table 14.
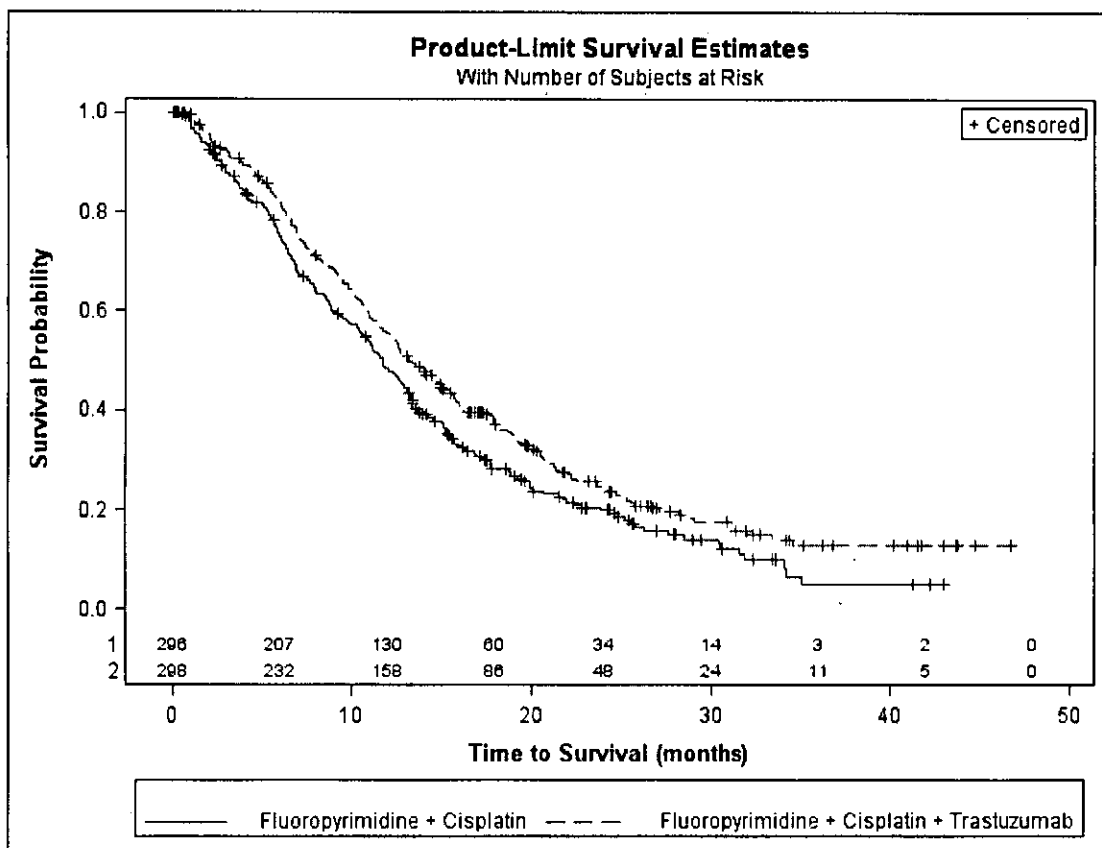
Figure 1. Updated Overall Survival in Patients with Metastatic Gastric Cancer.

Table 14. Overall Survival in ITT Population

	FC Arm N= 296	FC + H Arm N=298
Final Overall Survival		
No. Deaths (%)	184 (62.2%)	167 (56.0%)
Median	11.0	13.5
95% CI (mos.)	(9.4, 12.5)	(11.7, 15.7)
Hazard Ratio	0.73	
95% CI	(0.60, 0.91)	
p-value*, two-sided	0.0038	
Updated Overall Survival		
No. Deaths (%)	227 (76.7%)	221 (74.2%)
Median	11.7	13.1
95% CI (mos.)	(10.3, 13.0)	(11.9, 15.1)
Hazard Ratio	0.80	
95% CI	(0.67, 0.97)	

* Comparing with the nominal significance level of 0.0193

An exploratory analysis of OS in patients based on gene amplification (FISH) and protein-overexpression (IHC) testing is summarized in Table 15.

Table 15. Exploratory Analyses by HER2 Status Using the Updated Overall Survival Results.

	FC N=296[a]	FC+H N=298[b]
FISH+ / IHC 0, 1+ subgroup (N=133) No. Deaths / n (%) Median OS Duration (mos.) 95% CI (mos.)	57/71 (80.3%) 8.8 (6.4, 11.7)	56/62 (90.3%) 8.3 (6.2, 10.7)
Hazard ratio (95% CI)	1.33 (092, 1.92)	
FISH+ / IHC2+ subgroup (N=160) No. Deaths / n (%) Median OS Duration (mos.) 95% CI (mos.)	65/80 (81%) 10.8 (6.8, 12.8)	64/80 (80%) 12.3 (9.5, 15.7)
Hazard ratio (95% CI)	0.78 (0.55, 1.10)	
FISH+ or FISH-/IHC3+[c] subgroup (N=294) No. Deaths / n (%) Median OS Duration (mos.) 95% CI (mos.)	104/143 (73%) 13.2 (11.5. 15.2)	96/151 (64%) 18.0 (15.5, 21.2)
Hazard ratio (95% CI)	0.66 (0.5, 0.87)	

Median survival was estimated from Kaplan-Meier curves.

[a] Two patients on FC arm who were FISH+ but IHC status unknown were excluded from the analyses.

[b] Five patients on Herceptin® arm who were FISH+ but IHC status unknown were excluded from the analyses.

[c] Includes 6 patients on chemotherapy arm, 10 patients on Herceptin® arm with FISH-, IHC3+ and 8 patients on chemotherapy arm, 8 patients on Herceptin® arm with FISH status unknown, IHC3+.

XI.   PANEL MEETING RECOMMENDATION AND FDA'S POST-PANEL ACTION

In accordance with the provisions of section 515(c)(2) of the act as amended by the Safe Medical Devices Act of 1990, this PMA was not referred to the Immunology Devices Panel, an FDA advisory committee, for review and recommendation because the information in the PMA substantially duplicates information previously reviewed by this panel.

XII.   CONCLUSION.S DRAWN FROM PRECLINICAL AND CLINICAL STUDIES

Day-to-day agreement showed very good reproducibility with overall HER2 status agreements ranging from 83.1% to 98.3% in the 60 comparisons. The vast majority of comparisons (47 of 60) resulted in overall agreements at 90.0% or above, and in 20 of the 60 comparisons the overall agreement was found at 95.0% or above. Fisher's exact test revealed that the observed results were not different between days.

Results obtained by the three first observers at the three sites (site-to-site agreement) showed agreement in HER2 status in the range 68.3% to 90.0%. The average overall agreements for the three possible site comparisons were found at 82.7%, 75.0% and

88.0%. According to Fisher's exact test the results obtained were not different between sites. When comparing HercepTest™ results obtained by automated and manual platforms (automated-to-manual agreement) good HER2 status agreement was observed. From the four preplanned comparisons the overall agreement was between 86.7% and 93.3% with average agreements for observer one at 90.0% and 90.9% for observer two.

HER2 status agreement between observers at each site (observer-to-observer agreement) was found to be 88.0%, 83.6% and 81.0% for Sites 1, 2, and 3, respectively. While the analytical performance for HercepTest™ is not optimal relative to performance for breast cancer, based on statistical analysis the agreement within and between observers at each site appears to be good, however differences between sites and observers between each site were observed. There were also difference noted between the observers and sites based on specimen type (biopsy vs. surgical resection) and tissue source (stomach vs. GEJ). These discordances may be due to specimen heterogeneity. An evaluation of 40 specimens from the clinical trial showed heterogeneity in 37.5% (15/40) with the more heterogeneous being found in the GEJ. As seen with the specimen heterogeneity studies performed at Targos and Dako, the size of the biopsy and the amount of tumor present can directly influence the call for the specimen as fewer biopsy specimens may not be representative of the actual HER2 status of the tumor.

In the effort to minimize the effect of heterogeneity and variability, additional recommendations for performance and interpretation of HercepTest™ staining were added to the assay's package insert, evaluation guide, and training materials.

A. Safety Conclusions

As a diagnostic test, the HercepTest™ assay involves testing on formalin-fixed, paraffin embedded human breast cancer tissue sections. These tissue sections are routinely removed for breast cancer diagnosis. The test, therefore, presents no additional safety hazard to the patient being tested.

B. Benefit/Risk

The analytical performance for HercepTest™ is not optimal relative to performance for breast cancer, however based on statistical analysis the agreement within and between observers at each site appears to be satisfactory.

Nearly all patients enrolled in the trial had tumors which were gene amplified (FISH+) and only 16 patients where either negative for FISH or FISH status was unknown. Patients whose tumors were gene amplified but not HER2 protein over-expressing (i.e., FISH +/IHC 0, 1+) were shown to not benefit in an exploratory analysis but those whose tumors were gene amplified but demonstrated weak to moderately (equivocal) HER2 protein over-expression (i.e., FISH +/IHC 2+) did appear to benefit, though not as much as those whose tumors were gene amplified and HER2 protein over-expressing (i.e., FISH +/ IHC 3+), as shown in Table 15. Because there were no patients whose tumors were not gene amplified but HER2 protein

weakly to strongly over-expressing [FISH(-)/IHC 2+] and an insufficient number of cases whose tumors were FISH(-)/IHC 3+ to allow for any estimate of efficacy, it is therefore unclear if patients whose tumors are not *HER2* gene amplified but Her2 protein over-expressing (i.e., IHC 2+ or 3+) will benefit from Herceptin® treatment.

Based on the preclinical and clinical analyses, patients' HER status should not be determined using a single method, and unlike with breast cancer testing, reflex testing for both IHC 2+ and 3+ for gene amplification status should be considered.

C. Overall Conclusions

Based on the preclinical and clinical data, FDA concludes that there is reasonable assurance of safety and effectiveness of this device for use in the assessment of Her2 protein-overexpression in conjunction with gene amplification testing is sufficient to effectively identify the appropriate patients to be considered for Herceptin® therapy.

XIII. CDRH DECISION

CDRH issued an approval order on October 20, 2010 concurrently with CDER's approval for the new indication for Herceptin® (trastuzumab) for use with metastatic gastric or gastroesophageal adenocarcinoma. The applicant's manufacturing facilities did not require additional inspection as this product is currently approved for marketing for another indication (breast cancer) and were found to be in compliance with the device Quality System (QS) regulation (21 CFR 820).

XIV. APPROVAL SPECIFICATIONS

Directions for use: See device labeling.

Hazards to Health from Use of the Device: See Indications, Contraindications, Warnings, Precautions, and Adverse Events in the device labeling.

Post-approval Requirements and Restrictions: See approval order.

XV. REFERENCES

1. Hofmann M, Stoss O, Shi D, *et al.* Assessment of a HER2 scoring system for gastric cancer: results from a validation study. *Histopathology* 2008;52(7):797-805.

27