

SUMMARY OF SAFETY AND EFFECTIVENESS DATA (SSED)

I. GENERAL INFORMATION

Device Generic Name: Device for Detection of HER-2/*neu* Gene Amplification using Fluorescent In Situ Hybridization (FISH)

Device Trade Name: *HER2* FISH pharmDx™ Kit

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: P040005/S005

Date of FDA Notice of Approval: October 20, 2010

Expedited: Not applicable.

The original PMA (P040005) for *HER2* FISH pharmDx™ Kit was approved on 05/03/2005 for breast cancer patients and is indicated as an aid in the assessment of patients for whom Herceptin™ (trastuzumab) treatment is being considered. The SSED to support the already approved indication is available on the CDRH website and is incorporated by reference here. The current supplement was submitted to expand the indication for the *HER2* FISH pharmDx™ Kit to metastatic gastric cancer patients.

II. INDICATIONS FOR USE

For in vitro diagnostic use.

HER2 FISH pharmDx™ Kit is a direct fluorescence in situ hybridization (FISH) assay designed to quantitatively determine HER2 gene amplification in formalin-fixed, paraffin-embedded (FFPE) breast cancer tissue specimens and FFPE specimens from patients with metastatic gastric or gastroesophageal junction adenocarcinoma.

HER2 FISH pharmDx™ Kit 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 *HER2* FISH pharmDx™ Kit labeling.

V. **DEVICE DESCRIPTION**

The device, *HER2* FISH pharmDx™ Kit does not differ from the previously approved device as described in the original device description in regard to manufacturing, quality control.

The fundamental differences between the kits approved for breast cancer vs. gastric cancer are procedural and recommendations with regard to specimen preparation, specimen evaluation, signal enumeration, and disregarding signal generated from bacterial DNA.² 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 to preserve the tissue for FISH analysis. Standard methods of tissue processing for immunohistochemical staining should be used for all specimens. When testing small biopsy specimens, ascertain intact tumor morphology and the presence of sufficient nuclei for enumeration. If *HER2* FISH analysis is performed on a biopsy specimen, multiple (7-8) evaluable biopsies from different regions of the tumor should be analyzed to ensure reliable determination of *HER2* status.

Assessable tissue

Locate the tumor within the context of the H&E stained slide and evaluate the same area on the FISH stained slide (in the DAPI filter). Only specimens from patients with adenocarcinoma of the stomach, including gastroesophageal junction should be analyzed. In cases with intestinal metaplasia and adenocarcinoma in the same specimen, only the carcinoma component should be scored. Avoid areas of heavy inflammation, necrosis and areas where the nuclear borders are ambiguous. Do not include nuclei that require subjective judgment. Do not include nuclei with weak signal intensity and non-specific or high background.

Begin with a microscope evaluation of the complete FISH stained section and the area assigned on the H&E section, respectively. Before enumeration of the FISH stained section, note the overall signal distribution (homogenous or heterogeneous) on the signal enumeration sheet. In case of heterogeneous distribution, note whether focal amplification or single cell amplification (mosaic) is present.

1) Homogenous signal distribution

In case the signal distribution is homogenous, enumerate the number of chromosome centromeres (green signals) and the number of *HER2* genes (red signals) respectively, from 20 cells in 1-2 representative tumor areas.

2) Heterogeneous signal distribution

In case the signal distribution is heterogeneous, enumerate a total of 20 cells from selected areas as specified below:

- A) If focal amplification exists, areas with amplified cells should be selected
- B) If mosaic distribution or amplified, polysomal and disomal cells are present, count in areas with amplified cells. Within these areas, not only amplified cells but also adjacent non-amplified cells should be counted for a total of 20 cells.

If possible, do not select overlapping areas.

Disregard staining of bacterial DNA

A number of specialized cells (mast cells and macrophages), present interspersed in the gastric tissue, exhibit a high level of staining by the *HER2* probe due to presence of bacterial DNA. This results in highly red fluorescent cells that are clearly distinct from tumor cells with high *HER2* gene amplification.

Signal enumeration:

- When an area has been selected for signal evaluation, begin analysis in one of the 20 adjacent chosen nuclei and then count in a cell-by-cell fashion only leaving out nuclei that do not meet the quality criteria.
- The distance has to be at least equal to the diameter of one normal-sized signal in order to count two individual signals. When the distance between two signals is less than the diameter of a signal it is counted as one.
- If the ratio is borderline (1.8-2.2), count an additional 40 nuclei and calculate the ratio for the 40 nuclei. If the enumeration continues to be borderline, the result of the second evaluation is valid. If available, the immunohistochemical staining of *HER2* should be included for better orientation during the second enumeration.

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

HER2 FISH pharmDx™ Kit for the extended indication in gastric cancer has not been marketed within the United States. *HER2* FISH pharmDx™ Kit 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

HER2 FISH pharmDx™ Kit is intended for in vitro diagnostic use only. As with any in vitro diagnostic test, the potential risks are associated with incorrect result interpretations. A false positive test result would likely assign patients to receive a potentially ineffective therapy, possibly exposing the patient to serious side effects and, in rare cases, death. Alternatively, a false negative test result may exclude a patient who might benefit from therapy, potentially resulting in a poor outcome. 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

Analytical performance testing was performed to evaluate the safety and effectiveness of HER2 FISH pharmDx™ Kit on gastric cancer tissue. The validation included analytical sensitivity, analytical specificity, hybridization efficiency, robustness and repeatability testing. Since reagent formulation, kit configuration and staining procedure remained unchanged, the analytical performance of HER2 FISH pharmDx™ kit proved to be comparable to studies performed on breast cancer tissue. The analytical performance validation studies concluded that there are no differences in HER2 FISH pharmDx™ Kit's performance and robustness when applied to gastric cancer tissue when compared to the performance on breast cancer tissue.

Analytical Validation Specimens

The materials for the validation study were formalin-fixed, paraffin-embedded (FFPE) human adenocarcinoma tissue specimens of the stomach or gastro-esophageal junction. The specimens included in the study were purchased leftover specimens and not part of the ToGA trial. The specimens were selected to include both resection and biopsy (original and biopsy substitutes) specimens representing different levels of HER2 gene amplification as well as homo- and heterogenic signal distribution. A biopsy substitute consists of two punches (3 mm) from a resection specimen. Selected tissue blocks representing the tissue specimens specified above were at study start cut into serial sections of approximately 4 µM and stored in the dark at 2-8°C. Stability for up to 31 months has been established for cut tissue sections of breast specimens for use in FISH analysis. Same stability is assumed for gastric cancer tissue specimens and will be verified in an ongoing stability study.

A series of analytical laboratory studies were performed and a summary is presented in this section.

A. Laboratory Studies

1. Non-Clinical Studies - Internal

a. Analytical Sensitivity on gastric cancer specimens

The analytical sensitivity of the *HER2/CEN-17* Probe Mix when used on gastric cancer tissue was investigated using 18 gastric cancer adenocarcinoma specimens. The ratio between the number of *HER2* signals and CEN-17 signals was calculated based on a counting of 20 nuclei from normal cells surrounding the tumor. The *HER2/CEN-17* ratio was scored between 0.91 and 1.09.

b. Analytical Specificity

End-sequencing and cross-hybridization is not tissue dependent and was not re-evaluated since the *HER2/CEN-17* probe mix configuration and sequence is unchanged. The data for end-sequencing and cross-hybridization (analytical specificity) was provided in the original PMA SSED.

c. Hybridization efficiency on gastric cancer specimens

Hybridization efficiency of HER2 FISH pharmDx™ Kit when used on gastric cancer tissue was investigated as part of the (external) reproducibility study. From the total 360 formalin-fixed, paraffin-embedded tissue sections tested at

the three study sites 358 could be enumerated in accordance with product guidelines. Thus, the hybridization efficiency was 99.4%.

d. Robustness on gastric cancer specimens

The robustness test of *HER2* FISH pharmDx™ Kit when staining gastric cancer tissue was designed similar to the robustness test on breast cancer tissue as presented in the approved PMA (P040005) for breast cancer. The Tests used two different FFPE human gastric cancer specimens with and without *HER2* gene amplification. All tests are performed in duplicates. Interpretation of staining (signal intensity, background, and tissue structures) was performed using intervals covering 0.5 grades for signal intensity and tissue structure. When a result was listed as an interval the lower limit of the interval was used to assess whether a test was within the acceptance criteria. Signal intensity, background, and tissue structures were compared to the reference conditions for each of the parameters tested. The reference is defined by the recommendation for the individual parameter stated in the package insert for *HER2* FISH pharmDx. In each test only one parameter was changed and all other parameters were kept at standard condition. Testing included varying pre-treatment time and temperature, pepsin incubation time, denaturation temperature, hybridization time and temperature, and stringent wash time, temperature and buffer concentration.

No significant difference in results was observed at the following experimental conditions:

- Pre-treatment for 7, 9, 10, 11, and 13 minutes at 95-97 °C
- Pre-treatment at 89, 92 and 95-97 °C for 10 minutes
- Pepsin incubation times of 2, 2½, 3 and 4 minutes at 37 °C
- Denaturation temperatures of 72, 77, 82, 87 and 92 °C for 5 minutes
- Hybridization times 10, 12 and 14 hours
- Hybridization temperatures at 40, 45 and 50 °C
- Stringent wash for 5, 10 and 15 minutes at 65 °C
- Stringent wash at 60, 65 and 70 °C for 10 minutes
- Stringent wash buffer concentration at 1:10, 1:15, 1:20, 1:30 and 1:40

Note: It is recommended to adhere to the time and temperatures indicated in the staining procedure provided in the package insert.

e. Precision (Repeatability)

Repeatability (Intra-run)

The repeatability of the *HER2*/CEN-17 ratio was investigated with *HER2* FISH pharmDx™ Kit using 3 consecutive sections from 9 different gastric cancer adenocarcinoma specimens (3 non-amplified, 3 IHC *HER2* 2+, and 3 *HER2* amplified) including stomach and GEJ specimen. Each specimen was tested in triplicate in the same staining run. The coefficient of variance was found to be 1-5%.

Table 1. Results of repeatability – Intra-run

Sample	Mean <i>HER2</i> /CEN-17 ratio	St. Dev.	%CV
58880	1.12	0.05	5.00
54201	1.37	0.04	3.00
52134	1.41	0.03	2.00
58882	1.21	0.01	1.00
59242	1.11	0.03	3.00
58883	2.82	0.05	2.00
58888	6.61	0.19	3.00
58887	3.50	0.06	2.00
58918	6.30	0.11	2.00

Repeatability - Section thickness

Repeatability on consecutive sections of gastric adenocarcinoma specimens (IHC *HER2* 2+) with different thickness (2-7 µm) was tested with the *HER2* FISH pharmDx™ Kit. All samples were evaluated although 2 µm and 7 µm were over- and under digested respectively. As stated in the package insert for *HER2* FISH, Pepsin incubation time should be adjusted to the thickness of the specimen as well as the fixation history. Similar to HercepTest™ There is a minor trend toward lower count for both *HER2* and CEN-17 signals with decreasing section thickness; however the gene status (amplified/non-amplified) of the specimen is unchanged. The study showed that section thickness does not influence the *HER2*/CEN-17. This study supports our recommendation in the package insert of section thickness from 3-6 µm for *HER2* FISH analysis on gastric tissue FFPE specimens. The coefficient of variance of the *HER2*/CEN-17 ratio in this study was found to be 2-6% i.e. in the same range as for tissue of equal thickness.

f. Reproducibility – lot-to-lot

To demonstrate the degree of agreement between repeated measurements of *HER2*/CEN-17 ratio carried out using three different production lots of *HER2* FISH using 9 different FFPE human gastric cancer specimens (3 non-amplified, 3 IHC *HER2* 2+ and 3 *HER2* amplified) including stomach and GEJ specimens, resection specimens and biopsy substitutes. Each specimen was tested using three production lots of *HER2* FISH and the *HER2*/CEN-17 ratio were calculated together with the CV for all repeats.

Table 2. Results of lot-to-lot reproducibility

Sample	<i>HER2/CEN-17</i> ratio mean	SD	%CV
58882	1.20	0.05	4.0
58887	3.05	0.05	2.0
54201	1.26	0.07	5.0
58883	2.87	0.10	4.0
58880	1.06	0.06	5.0
59242	1.03	0.05	5.0
52134	1.32	0.13	10.0
58888	4.78	0.52	11.0
58918	5.10	0.23	5.0

g. Observer-to-observer (on same slide)

A reproducibility study was conducted using 9 different FFPE human gastric cancer specimens. The specimens consisted of 3 non-amplified, 3 IHC *HER2* 2+, and 3 *HER2* amplified cases, which included stomach and GEJ specimens, resection specimens and biopsy substitutes to demonstrate the degree of agreement between evaluations of *HER2/CEN-17* ratio carried out by different observers on the same slides. The samples were blinded before evaluation.

Each specimen was evaluated by three independent observers and the *HER2/CEN-17* ratio was calculated together with the CV for all observations. A summary of the results are presented in Table 32 and a graphic illustration of the results are presented in Figure 4.

Table 3. Mean observer-to-observer results.

Sample	<i>HER2/CEN-17</i> ratio mean	SD	%CV
1	1.06	0.08	7.0
2	1.20	0.12	10.0
3	1.23	0.15	12.0
4	1.17	0.02	2.0
5	1.06	0.03	3.0
6	2.96	0.21	7.0
7	6.03	0.70	12.0
8	3.59	0.53	15.0
9	7.24	2.59	36.0

The variance between observers was higher (%CV= 2 – 15%) than the variance obtained in the repeatability (intra-run) %CV=1 – 5%. In one of the amplified specimens one observer counted a much higher number of red *HER2* signals as well as a lower number of green *CEN-17* signals. This results in a *HER2/CEN-17* ratio much higher than the one obtained by the two other observers and a %CV = 36% which is above what is accepted. However, in general, a higher degree of variance is to be expected in highly amplified specimens and this does not include a risk for wrong diagnosis of the patient.

h. Reproducibility – Tissue type

The clinical study BO18255 was performed on two tissue types, namely stomach and GEJ cancer consisting of both resection and biopsy specimens. Both tissue types are represented in both the internal and external validation studies in all categories (non-amplified, IHC HER2 2+ and amplified).

Reproducibility of *HER2* gene status by FISH was assessed on original biopsy specimens to show that it is reproducible from day-to-day and from observer-to-observer. Due to lack of sufficient material in an original biopsy specimen the inclusion of that specimen type is addressed in this study rather than the external reproducibility study. Six original biopsy specimens representing 3 stomach and 3 GEJ cancer specimens were stained on three non-consecutive days and scored by three independent observers. *HER2* gene status was performed according to the standard staining protocol. Each of the original biopsy specimens were stained three times in three independent staining runs. For each staining run each specimen was evaluated by three independent observers and the *HER2/CEN-17* ratio was calculated together with the CV for all observations.

The coefficient of variance (day-to-day and observer-to-observer) is higher for the original biopsy specimens than what has been observed in intra-run (1 – 5%) and observer-to-observer (2 – 22%) that is performed on resection specimens and biopsy substitutes. The variability in both studies was higher than expected variability was noted in both studies, although the variability could be due to the amplification pattern (focal vs. mosaic), tissue heterogeneity, or areas on the slide enumerated by the readers.

2. Non-Clinical Studies – External

External non-clinical studies consisted of a three site reproducibility and precision study performed at three labs trained and experienced in the use of each of the assays. During staining and enumeration of the FISH slides the working procedure included in *HER2* FISH pharmDx™ Kit was followed. *HER2* gene status was defined as non-amplified when the *HER2/CEN-17* ratio is < 2.0 and as amplified when the *HER2/CEN-17* ratio is ≥ 2.0.

a. Reproducibility Study

The reproducibility study was a three-site, blinded study using 4 µm sections of resection and biopsy substitute specimens of formalin-fixed paraffin-embedded (FFPE) human gastric adenocarcinoma tissues from the stomach or GEJ. The 24 specimens were equally distributed in the following three categories: non-amplified, IHC HER2 2+ (selected using HercepTest™) and amplified. At each of three study sites sections from 24 different FFPE specimens were stained in five separate runs on five non-consecutive days. The specimens represented surgical resections (70%) and biopsies (30%) with an equal number of non-amplified, IHC 2+ (determined by HercepTest™), and amplified specimens. The total number of sections stained at each study site was 120. Signal enumeration was performed on these 120 slides by two independent observers at each study site (observer one and observer two).

The average %CV for the *HER2*/CEN-17 ratio determined for each observer from the eight specimens in each category (five observations per block) is seen in Table 4. The overall average CVs for each observer were 5.4%, 3.8%, 12.0%, 11.9%, 4.7% and 24.4%, respectively. The average CVs determined in each category for all combined observations (days, sites, observers) were 22.8%, 16.5% and 25.2% in the non-amplified, IHC 2+, and amplified category, respectively.

Results - descriptive

Day-to-day %CVs for the first observers at the three sites were found in the range of 1.4% to 19.1% for the non-amplified category, in the range of 1.4% to 24.1% for the IHC 2+ category and in the range of 2.2% to 44.5% for the amplified category. For the second observers at the three sites day-to-day CVs were found in the range of 2.3% to 85.8% for the non-amplified category, in the range of 1.8% to 20.8% for the IHC 2+ category and in the range of 0.6% to 58.2% for the amplified category. For the first observers the average descriptive day-to-day CVs for all specimens was close to 5% for sites 1 and 3 and 12% for site 2. For the second observers the average descriptive day-to-day CVs for all specimens was 4%, 12%, and 24% for the sites 1, 2, and 3, respectively.

Table 4. Mean and CV (%) of *HER2*/CEN-17 ratios obtained.

Tissue block	<i>HER2</i> /CEN-17 ratio		Tissue block	<i>HER2</i> /CEN-17 ratio		Tissue block	<i>HER2</i> /CEN-17 ratio	
	Non-amplified			IHC 2+			Amplified	
	Mean	CV (%)		Mean	CV (%)		Mean	CV (%)
54210	1.27	15.04	53116	3.04	18.44	59248	21.07	25.48
54220	1.28	16.89	53832	1.13	11.71	59249	14.70	20.73
58752	1.39	20.42	58882	1.21	14.98	59257	9.89	35.33
59252	1.09	8.15	58892	1.12	13.97	59269	8.73	21.62
59259	1.20	28.99	59239	2.21	27.59	59272	11.31	23.98
59262	1.23	16.23	59241	1.14	10.97	59300	6.67	19.85
59297	1.66	65.46	59246	2.85	17.76	59302	7.80	33.83
59314	1.14	11.41	59254	1.32	16.31	59304	22.96	20.57

Each mean and CV is derived from the total of 30 enumerations made by the six different observers on five different days.

When combining all observations made on a tissue block, the mean and %CV for the 30 enumerations (five days by six observers) were as depicted in Table 4. From these observations average descriptive CVs representing the total variation (day, site, and observer variation) in the three categories was calculated at 22.8%, 16.5%, and 25.2% for non-amplified, IHC 2+, and amplified categories, respectively.

Site-to-site

Table 5A and 5B that compare the relevant counts of *HER2* signals from the amplified specimens between sites. As can be seen from Table 5B, for each of

the amplified specimens there seem to be a clear tendency that the mean *HER2* signal (mean of total *HER2* signals from 20 nuclei for a particular specimen as obtained during five runs by two observers) for Site 1 is elevated when compared to both Sites 2 and 3. In Table 5B a mean total (20 nuclei) *HER2* signal number for all specimens stratified by site (8 amplified specimens, 2 observers, 5 runs) are presented summarizing this observation.

Table 5A. Mean total *HER2* signals in 20 nuclei stratified by specimen.

Category	Block	Mean <i>HER2</i> signal number (in 20 nuclei)			N*
		Site 1	Site 2	Site 3	
Amplified	59248	1077	820	685	30
	59249	635	572	490	30
	59257	413	254	283	30
	59269	436	328	396	30
	59272	483	448	331	30
	59300	369	318	272	30
	59302	583	430	292	30
	59304	1122	1119	735	30

Values represent the mean total *HER2* counts counted in 20 nuclei for 5 runs, two observers, therefore each specimen has been counted 10 times at each of the three sites (n=30).

Table 5B. Total mean *HER2* signals stratified by site.

Category	Site	Mean <i>HER2</i> signal	N	SD
Amplified	Site 1	639.56	80	284.239
	Site 2	536.24	80	300.519
	Site 3	435.42	80	210.552
Total		537.07	240	279.617

Values represent the mean of the total counted *HER2* signals in 20 nuclei for the eight specimens at each site (8 specimens, 5 runs and 2 observers; n=80).

Observer-to-observer

HER2 status overall agreements stratified by specimens and observers are shown in Table 6. The mean ratio from the five runs for each specimen (n=24) were translated to *HER2* status and subsequently, 2x2 cross-tabulations for *HER2* status (non-amplified, amplified) were performed. Overall percent agreements as well as the lower and upper 95% confidence limits based on the binomial distribution are shown in the table below.

Table 6. *HER2* status overall agreements stratified by specimens.

Site and Observer Comparison	Overall Agreement (%)	95% CI Lower Limit	95% CI Upper Limit
Site 2, obs 1 vs. Site 1, obs 1	95.8	82.1	99.5
Site 2, obs 1 vs. Site 1, obs 2	95.8	82.1	99.5

Site 2, obs 2 vs. Site 1, obs 1	91.7	75.9	98.2
Site 2, obs 2 vs. Site 1, obs 2	91.7	75.9	98.2
Site 2, obs 1 vs. Site 3, obs 1	95.8	82.1	99.5
Site 2, obs 1 vs. Site 3, obs 2	100.0	90.2	100.0
Site 2, obs 2 vs. Site 3, obs 1	91.7	75.9	98.2
Site 2, obs 2 vs. Site 3, obs 2	95.8	82.1	99.5
Site 3, obs 1 vs. Site 1, obs 1	100.0	90.2	100.0
Site 3, obs 1 vs. Site 1, obs 2	100.0	90.2	100.0
Site 3, obs 2 vs. Site 1, obs 1	95.8	82.1	99.5
Site 3, obs 2 vs. Site 1, obs 2	95.8	82.1	99.5
Site 2, obs 1 vs. Site 2, obs 2	95.8	82.1	99.5
Site 3, obs 1 vs. Site 3, obs 2	95.8	82.1	99.5
Site 1, obs 1 vs. Site 1, obs 2	100.0	90.2	100.0

Obs – Observer

Agreement calculations stratified by specimen type and HER2 status was determined from the mean *HER2/CEN-17* ratios as described above and overall agreements calculated for each specimen type (biopsy; n=7 and surgical; n=17) from 2x2 cross-tabulations. Results of the agreement calculations are tabulated in Table 7. It should be noted that the wide 95% confidence intervals are caused by the low number of specimens represented in each layer when split into biopsies and surgical resections.

The tendency to a lower overall agreement for biopsies compared to surgical specimens presented in Table 7, which should be interpreted with caution, is caused by discordance for two of the biopsy specimens.

One of the two biopsies (block 59239) which was discordant was from a heterogeneous, focally amplified specimen with a *HER2/CEN-17* ratio in the borderline area (overall ratio is 2.21). For biopsy specimens it is a risk that the observer simply does not identify all cores on a slide and, therefore, does not find the correct tumor area to enumerate. To minimize this risk recommendations are given in Instructions for Use to be careful in identifying all tissue cores when analyzing biopsy specimens (i.e. inspection of the H&E stained section to reveal all tissue cores available).

The second biopsy showing discordance is based on one outlier reading for observer 2 at Site 2 (block 59297). By examination of raw data from the study it appeared that the *HER2/CEN-17* ratio was found at 7.11 for this block in run 1 and at 1.71, 1.73, 1.73 and 1.75 in runs 2, 3, 4, and 5 from this site and observer, respectively. Furthermore, the ratios obtained for other readings of this block were in the range 0.98 – 2.64. This specific reading of 7.11 results in a mean *HER2/CEN-17* ratio at 2.86, i.e. amplified for this site and observer.

Agreement calculations stratified by tissue type HER2 status was determined from the mean *HER2/CEN-17* ratios as described above and overall agreements calculated for each tissue type (GEJ; n=6 and stomach; n=18) from 2x2 cross-tabulations. Results of the agreement calculations are

tabulated in Table 8. It should be noted that the wide 95% confidence intervals are caused by the low number of specimens represented in each layer when split into tissue types.

Table 7. HER2 status overall agreements stratified by specimen type.

	Site and Observer Comparison	Overall Agreement (%)	95% CI Lower Limit	95% CI Upper Limit
Biopsy (n=7)	Site 2, obs 1 vs. Site 1, obs 1	85.7	49.9	98.4
	Site 2, obs 1 vs. Site 1, obs 2	85.7	49.9	98.4
	Site 2, obs 2 vs. Site 1, obs 1	71.4	35.2	93.5
	Site 2, obs 2 vs. Site 1, obs 2	71.4	35.2	93.5
Surgical (n=17)	Site 2, obs 1 vs. Site 1, obs 1	100.0	86.5	100.0
	Site 2, obs 1 vs. Site 1, obs 2	100.0	86.5	100.0
	Site 2, obs 2 vs. Site 1, obs 1	100.0	86.5	100.0
	Site 2, obs 2 vs. Site 1, obs 2	100.0	86.5	100.0
Biopsy (n=7)	Site 2, obs 1 vs. Site 3, obs 1	85.7	49.9	98.4
	Site 2, obs 1 vs. Site 3, obs 2	100.0	70.8	100.0
	Site 2, obs 2 vs. Site 3, obs 1	71.4	35.2	93.5
	Site 2, obs 2 vs. Site 3, obs 2	85.7	49.9	98.4
Surgical (n=17)	Site 2, obs 1 vs. Site 3, obs 1	100.0	86.5	100.0
	Site 2, obs 1 vs. Site 3, obs 2	100.0	86.5	100.0
	Site 2, obs 2 vs. Site 3, obs 1	100.0	86.5	100.0
	Site 2, obs 2 vs. Site 3, obs 2	100.0	86.5	100.0
Biopsy (n=7)	Site 3, obs 1 vs. Site 1, obs 1	100.0	70.8	100.0
	Site 3, obs 1 vs. Site 1, obs 2	100.0	70.8	100.0
	Site 3, obs 2 vs. Site 1, obs 1	85.7	49.9	98.4
	Site 3, obs 2 vs. Site 1, obs 2	85.7	49.9	98.4
Surgical (n=17)	Site 3, obs 1 vs. Site 1, obs 1	100.0	86.5	100.0
	Site 3, obs 1 vs. Site 1, obs 2	100.0	86.5	100.0
	Site 3, obs 2 vs. Site 1, obs 1	100.0	86.5	100.0
	Site 3, obs 2 vs. Site 1, obs 2	100.0	86.5	100.0
Biopsy (n=7)	Site 2, obs 1 vs. Site 2, obs 2	85.7	49.9	98.4
	Site 3, obs 1 vs. Site 3, obs 2	85.7	49.9	98.4
	Site 1, obs 1 vs. Site 1, obs 2	100.0	70.8	100.0
Surgical (n=17)	Site 2, obs 1 vs. Site 2, obs 2	100.0	86.5	100.0
	Site 3, obs 1 vs. Site 3, obs 2	100.0	86.5	100.0
	Site 1, obs 1 vs. Site 1, obs 2	100.0	86.5	100.0

Obs – Observer

Table 8. HER2 status overall agreements stratified by tissue type.

	Site and Observer Comparison	Overall Agreement (%)	95% CI Lower Limit	95% CI Upper Limit
GEJ (n=6)	Site 2, obs 1 vs. Site 1, obs 1	83.3	44.2	98.1
	Site 2, obs 1 vs. Site 1, obs 2	83.3	44.2	98.1
	Site 2, obs 2 vs. Site 1, obs 1	83.3	44.2	98.1
	Site 2, obs 2 vs. Site 1, obs 2	83.3	44.2	98.1
Stomach (n=18)	Site 2, obs 1 vs. Site 1, obs 1	100.0	87.1	100.0
	Site 2, obs 1 vs. Site 1, obs 2	100.0	87.1	100.0
	Site 2, obs 2 vs. Site 1, obs 1	94.4	76.8	100.0
	Site 2, obs 2 vs. Site 1, obs 2	94.4	76.8	100.0
GEJ (n=6)	Site 2, obs 1 vs. Site 3, obs 1	83.3	44.2	98.1
	Site 2, obs 1 vs. Site 3, obs 2	100.0	67.0	100.0
	Site 2, obs 2 vs. Site 3, obs 1	83.3	44.2	98.1
	Site 2, obs 2 vs. Site 3, obs 2	100.0	67.0	100.0
Stomach (n=18)	Site 2, obs 1 vs. Site 3, obs 1	100.0	87.1	100.0
	Site 2, obs 1 vs. Site 3, obs 2	100.0	87.1	100.0
	Site 2, obs 2 vs. Site 3, obs 1	94.4	76.8	100.0
	Site 2, obs 2 vs. Site 3, obs 2	94.4	76.8	100.0
GEJ (n=6)	Site 3, obs 1 vs. Site 1, obs 1	100.0	67.0	100.0
	Site 3, obs 1 vs. Site 1, obs 2	100.0	67.0	100.0
	Site 3, obs 2 vs. Site 1, obs 1	83.3	44.2	98.1
	Site 3, obs 2 vs. Site 1, obs 2	83.3	44.2	98.1
Stomach (n=18)	Site 3, obs 1 vs. Site 1, obs 1	100.0	87.1	100.0
	Site 3, obs 1 vs. Site 1, obs 2	100.0	87.1	100.0
	Site 3, obs 2 vs. Site 1, obs 1	100.0	87.1	100.0
	Site 3, obs 2 vs. Site 1, obs 2	100.0	87.1	100.0
GEJ (n=6)	Site 2, obs 1 vs. Site 2, obs 2	100.0	67.0	100.0
	Site 3, obs 1 vs. Site 3, obs 2	83.3	44.2	98.1
	Site 1, obs 1 vs. Site 1, obs 2	100.0	67.0	100.0
Stomach (n=18)	Site 2, obs 1 vs. Site 2, obs 2	94.4	76.8	100.0
	Site 3, obs 1 vs. Site 3, obs 2	100.0	87.1	100.0
	Site 1, obs 1 vs. Site 1, obs 2	100.0	87.1	100.0

Obs – observer

A possible explanation as to why *HER2/CEN-17* ratios at Site 1 were found to be 12% and 17% higher than the ratios obtained at Site 2 and Site 3, respectively. An explanation of this observation could be due to observers at Site 1 estimated highly amplified clusters at higher numbers than the observers at Site 3 and Site 2.

Results – based on models (variance component and ANOVA)

Day-to-day

From variance component models applied to the logarithmically transformed *HER2/CEN-17* ratios it was found that the variation within the tissue block itself (the residual error component) accounted for the day-to-day variation observed, indicating the absence of day-to-day variation in this study. Differences were also analyzed using an ANOVA (analysis of variance) model. This analysis confirmed that no day-to-day variation could be found.

Site-to-site

Analysis of site-to-site variation using variance component model analyses showed that a modest effect of the site was present indicating that a minor site-to-site variation was found when compared to the variation within the tissue block itself.

ANOVA showed that significant site-to-site differences were present both for the first observers ($F=23.37$, $p<0.001$) and for the second observers ($F=14.05$, $p<0.0001$). Further analysis revealed that study Site 1 had 12% and 16% higher *HER2/CEN-17* ratios compared to study Site two for first and second observers, respectively, and furthermore, that study Site 1 had 16% and 17% higher *HER2/CEN-17* ratios compared to study Site 3 for first and second observers, respectively. No significant differences were observed between study sites two and three.

Observer-to-observer

To determine observer-to-observer differences within each of the three sites a statistical variance component model was applied on the logarithmically transformed *HER2/CEN-17* ratios. It was found that for the observers at study sites two and three, the variation (e.g., heterogeneity and/or presence of clusters) within the tissue block itself (the residual error component) accounted for the complete observer-to-observer variation. For study site one a small part of the variation could not be accounted for by the tissue block itself, indicating a minor observer-to-observer difference at site one. Using an ANOVA model it was also found that no observer-to-observer variation could be identified at sites two and three. However, at site one, results showed that the *HER2/CEN-17* ratios obtained by observer two were 3.7% higher than the results obtained by observer one. Due to the low analytical variability found at site one a similar difference would not be significant at the two other sites.

Conclusion

The average CVs obtained by the six observers when enumerating signals from five sections from each of 24 tissue blocks revealed low average CVs at 5%, 12%, 5%, 4%, and 12%, for five observers and a relatively high average CV at 24% for one observer.

Using different statistical models it was found that the variation associated with the specimen itself contributed to the vast majority of the day-to-day, site-to-site and observer-to-observer variation. However, minor but significant contributions to the total variation observed were identified and could be attributed to site or observer differences. This can be explained in that if the observers do not enumerate the same areas and/or nuclei, variability in results

is expected. Of greatest importance is that the final resulting call is accurate, which is observed to be the case. It was found that study site one had between 12% and 17% higher *HER/CEN-17* ratios compared to the two other sites. From a review of the raw data, the majority of the specimens exhibited cluster amplification, which is known to result in variability between observers due to the require the reader to estimate the number of signals represented by the clustered signal. Also a difference of 3.7% was found between the two observers from site one, whereas no difference was observed between sites two and three and between observers at these sites.

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 FISH specimens was provided and is summarized below.

a. Heterogeneity Study Performed at Targos Molecular Pathology GmbH

A heterogeneity assessment study was performed at the Targos central laboratory. The assessment was performed on selected IHC and FISH specimens from BO18255 trial. Only results of the assessment of FISH specimens were provided in the PMA supplement.

Table 9. Patient cases included in the study by FISH score and primary site

FISH status	Stomach	GEJ	Total
Amplified	11	13	24
Non-amplified	2	0	2
Total	13	13	26

The evaluated FISH slides from BO18255 trial were selected based on the following criteria: FISH specimens were selected based on age, i.e. how long slides were stored after staining was performed. Due to the high likelihood of signal fading over time only slides stained in 2008 or 2007 were selected for heterogeneity assessment. Selected cases were then evaluated by a pathologist at Targos to determine if the slide passes quality control. Only slides that passed quality control were evaluated for heterogeneity. FISH cases were

selected from patients with stomach or GEJ cancer in IHC 0, 1+, 2+, and 3+ categories. A total of 26 FISH cases were evaluated and summarized in the Table 9 above.

For each selected patient case the biopsies on the slide were evaluated and heterogeneity on slide was reported. For FISH, heterogeneity is defined as individual pieces having different HER2 status, i.e. some pieces are amplified while others are not, in a specimen. The number of evaluable biopsy pieces in each patient case varies due to the followings: 1) not all biopsy pieces were present on the FISH slide, 2) no relevant tissue was available in some of the biopsy pieces or 3) the biopsy pieces did not pass quality control criteria.

Results and Conclusions

The percent tumor heterogeneity was calculated for specimens from primary site of stomach and GEJ. For stomach 54% (7/13) and GEJ 23% (3/13) of cases demonstrated heterogeneity, yielding to an overall heterogeneity of 38.5% (10/26) for biopsy specimens analyzed in the study. The study indicates that one FISH-positive case reported in BO18255 trial (amplified with *HER2/CEN17* ratio of 2.05) became negative (non-amplified with *HER2/CEN17* ratio ranging from 1.26-1.4) when re-evaluated for heterogeneity analysis. This may have been due to signal fading over time as the case started as a borderline case.

The biopsy cases in BO18255 trial included 1 – 7 evaluable biopsy specimens. Heterogeneity observed on biopsy cases in this study relates to the heterogeneity at the gross tumor level (i.e. sampling from different locations in the tumor) and that both tumors from stomach and GEJ exhibited tumor heterogeneity (54% in stomach, 23% in GEJ, and 38.5% overall). Based on the findings it's concluded that that multiple biopsy pieces should be evaluated for reliable HER2 status determination due to heterogeneity issue.

b. Heterogeneity Study Performed at Dako Denmark

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.

The tumor area 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) to mimic biopsy samples. The study was performed using FFPE specimens from human gastric and GEJ cancer tissue specimens from 24 different patients (i.e. each slide is from a different patient). The specimens included in the study are not individually identifiable, and it is not possible to trace the identity of the patients. The specimens represent both stomach and GEJ in three categories: Non-amplified, IHC HER2 2+ and Amplified (see Table 10).

Table 10. Distribution of resection specimens tested in the study

<i>HER2</i> status	Stomach	GEJ	Total
Non-Amp	4	3	7
IHC <i>HER2</i> 2+	5	5	10
Amp	4	3	7
Total	13	11	24

The tumor area was divided into 2 mm x 2 mm squares and 9 squares (randomly selected) were evaluated for each specimen. The total number of squares covering the specimens ranged from 7-92. Unique random numbers within the number of squares possible were selected using the RAND function of Excel. In all specimens one or several squares representing the margin have been evaluated and no edge artifacts that influenced scoring were observed. To evaluate the reliability of a result based on a biopsy specimen the *HER2* gene status of the resection specimen was compared to the combined/final score of square 1-3, 1-6 and 1-9. The combined/final score for square 1-3, 1-6 and 1-9 is amplified if there are one or more amplified squares among the three, six and nine squares otherwise the combined/final score is non-amplified.

For each of the 24 specimens the status of the resection and each of the nine evaluated squares were reported. Nine of the specimens (9/24=37.5%) exhibited heterogeneity between different squares and these specimens were almost equally distributed between stomach (4) and GEJ (5). Interpretation of *HER2* FISH on stomach and GEJ specimens was performed as described in the package insert for *HER2* FISH on gastric cancer. The package insert includes a description on how to assess the specimen with regard to signal distribution and how to perform signal enumeration in the selected area. Interpretation was performed identically for the complete resection specimen and for the individual squares. Most of the specimens used in this study are heterogeneous, and represent all signal distribution categories (homogeneous, heterogeneous-focal and heterogeneous mosaic) and exhibit *HER2/CEN-17* ratios between 0.97-27.97 with many samples located close to the borderline range (1.8-2.2). The *HER2/CEN-17* ratio was transformed to a *HER2* status: *HER2/CEN-17* ≥ 2.0 is amplified and *HER2/CEN-17* < 2 is non-amplified.

Results

Of the 11 non-amplified specimens there was one specimen where the combined score of the squares deviated from the score of the resection specimen i.e. one square was scored as amplified. The squares in this specimen exhibited *HER2/CEN-17* ratios between 1.61 and 2.3 with 3 squares in the borderline range (40 additional nuclei were counted for these squares). A specimen with a *HER2/CEN-17* ratio in the borderline range (1.8 - 2.2) should always be interpreted with caution. All other resection specimens with non-amplified score exhibited no heterogeneity within the section (between squares).

Heterogeneity was demonstrated in eight of the thirteen amplified specimens, i.e. one or more of the squares were scored as non-amplified. Of those eight, three specimens had at least one non-amplified square, one with two non-amplified squares, and four showed at least three or more non-amplified squares. This was due to a heterogeneous-focal signal distribution. These specimens represent cases where the number of biopsy cores and scanning (for resection specimens) is most important since there is a risk of missing a focal amplified area in the tumor. Many of the non-amplified specimens were heterogeneous; however they exhibited mosaic signal distribution which only rarely leads to a *HER2*/CEN-17 ratio ≥ 2.0 . As expected some of the squares in the amplified specimens were scored as non-amplified.

To evaluate the reliability of a result based on a biopsy specimen the *HER2* gene status of the resection specimen was compared to the combined/final score of squares 1-3, 1-6, and 1-9. The percentage of specimens where the combined/final status of the 3, 6 or 9 squares evaluated is the same as the status for the resection specimen was calculated. Percentage of specimens with the combined score is identical to the score of the resection specimen was 96% for 3, 6 and 9 squares. The evaluation of the squares for the combined score was based on the evaluation of sequential squares and not potential randomization of squares evaluated.

A re-evaluation of the Dako heterogeneity study was performed to re-validate the results of the first study on the heterogeneous specimens (N=9). 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.

All squares in non-amplified resection specimens (non IHC 2+) were scored as non-amplified. This was expected since the scoring of resection specimens and biopsy cores are performed according to same scoring guide and scoring was performed in the most amplified area. To evaluate the reliability of a result based on a biopsy specimen the *HER2* gene status of the resection specimen was compared to the combined/final score of square 1-3, 1-6, and 1-9. In the first study of the 24 specimens analyzed, the only specimen where the combined score of three squares disagree with the score for the resection. The squares in this specimen exhibit *HER2*/CEN-17 ratios between 1.61 and 2.3 with 3 squares in the borderline range (40 additional nuclei were counted for these squares).

The nine specimens that showed heterogeneity between squares and/or resection specimen in the original study were re-evaluated. The re-evaluated specimens represent both stomach (4) and GEJ (5). Unique random numbers within the number of squares possible were selected using RAND function of Excel. In some cases the same squares which were selected previously were counted. When comparing the re-evaluation to the original study, all squares that have been analyzed both times have achieved the same *HER2* status.

Conclusion – Dako Heterogeneity studies

This study includes the reassessment of 9 heterogeneous specimens from the original study. A heterogeneous specimen in this context is defined as a specimen where the HER2 status (amplified or non-amplified) as determined in a resection specimen is different from individual areas in the specimen (squares). The re-evaluation showed that eight out of nine specimens (selected based on the original study) exhibited heterogeneity in HER2 status between squares and/or the entire resection specimen. This indicates that in gastric cancer both stomach and GEJ specimens demonstrate heterogeneity in HER2 amplification pattern and that several biopsies should be evaluated for each patient case to obtain a reliable HER2 result. With respect to the recommendation for analyzing multiple biopsies that should be included in the package insert for *HER2* FISH pharmDx™, the rationale is as follows:

- Heterogeneity in HER2 amplification is common in gastric cancer specimens and several biopsies should be evaluated for each patient case to ensure reliable determination of HER2 status
- Both stomach and GEJ specimens show heterogeneity
- In two instances in the original and three in the reassessment study, more than half of the nine squares (biopsies) were discordant with the resection specimen status, indicating more than six or five biopsy specimens, respectively, should be required to obtain a reliable result compared to the resection specimen.

Based on the results of the heterogeneity studies and a parallel study using HercepTest™, a recommendation that 7-8 biopsies should be evaluated from based on more conservative estimates may be necessary to achieve a better determination of the resection specimen's gene amplification status.

Although gastric cancer tissue is heterogeneous this study on 24 different gastric cancer resection specimens shows that the analysis of biopsy cores can provide reliable results when compared to analysis of the full resection specimen; however the number of biopsy cores evaluated can significantly impact the reliability of results. There is no evident effect of analyzing tissue at the margin of the specimen and also the number of tumor cells is not crucial as long as the biopsy core includes at least 20 assessable nuclei for enumeration.

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 summary of the clinical study is presented below.

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 (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 specimens or biopsies.

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 in the study, 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).

B. Accountability of PMA Cohort

A total of 594 patients were enrolled to the study, 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 full analysis population with advanced gastric cancer from the BO18255 study are shown in Table 11.

Table 11. 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%

^a Source: U.S. Census Bureau 2006 estimate

^b Source: SEER-17 (2004-2006) advanced gastric carcinoma population, defined as Stage IIIB/IV, based on the November 2008 submission.

^c Source: Clinical Study Report (BO18255), enrolled September 2005 to December 2008. The all-randomized population included all subjects 594 who were randomized to treatment in the study, regardless of whether they actually received any study treatment.

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 12. 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 12. Summary of Patient Demographic Data

	FC (n=296)	FC+H (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%)

The stratification factors were well-balanced between the treatment arms as shown in Table 13. 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 14. 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 13. Summary of Stratification Factors

	FC (n=296)	FC+H (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 14. 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 15.

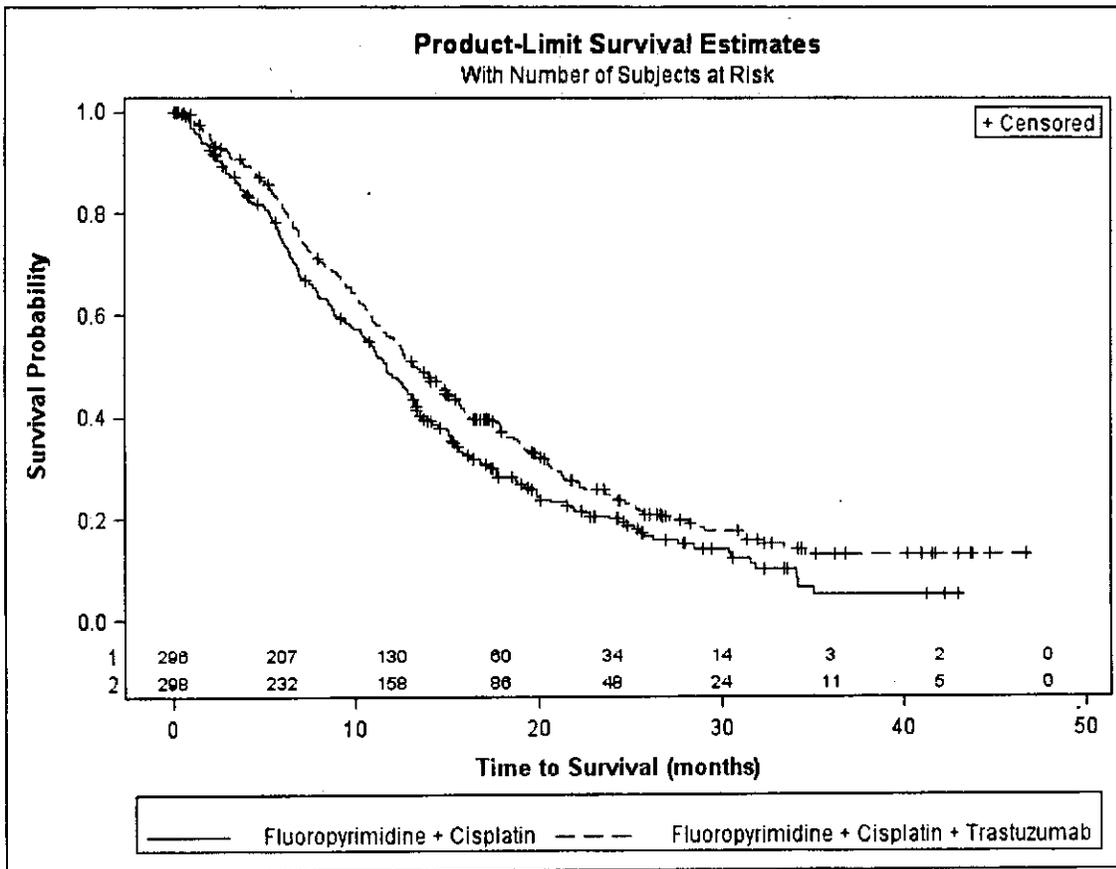


Figure 1. Updated Overall Survival in Patients with Metastatic Gastric Cancer.

Table 15. 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 16.

Table 16. 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 (%)	57/71 (80.3%)	56/62 (90.3%)
Median OS Duration (mos.)	8.8	8.3
95% CI (mos.)	(6.4, 11.7)	(6.2, 10.7)
Hazard ratio (95% CI)	1.33 (0.92, 1.92)	
FISH+ / IHC2+ subgroup (N=160)		
No. Deaths / n (%)	65/80 (81%)	64/80 (80%)
Median OS Duration (mos.)	10.8	12.3
95% CI (mos.)	(6.8, 12.8)	(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 (%)	104/143 (73%)	96/151 (64%)
Median OS Duration (mos.)	13.2	18.0
95% CI (mos.)	(11.5, 15.2)	(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. CONCLUSIONS DRAWN FROM PRECLINICAL AND CLINICAL STUDIES

Day-to-day agreement showed very good reproducibility with overall *HER2* gene amplification agreements. From analysis by several methods, only a small amount of variability in the reproducibility results could be contributed by variation between sites and the vast majority of variation was accounted for by the specimen itself which in turn affected some observer-to-observer results, contributing to higher than normal %CVs. The CV estimates determined using variance component models indicated that variation were lower in the IHC 2+ category of specimens. This demonstrated the need for and adherence to additional guidelines for sample evaluation and signal enumeration which are indicted in the package insert and user interpretation guide.

The differences in gastric cancer versus breast cancer histopathology, which include incomplete membrane staining (by IHC) and degree of heterogeneous expression of *HER2* in gastric cancers observed in the clinical study demonstrate that gene amplification and protein overexpression are not as well correlated as with breast cancer. For these reasons, users are not advised to rely on a single testing method to assess *HER2* status.

The safety of the therapeutic intervention with chemotherapy and trastuzumab [(FC+H) or (FC)] will not be addressed in the SSED for *HER2* FISH pharmDx™ Kit. Based on the results of the BO18255 study, the *HER2* FISH pharmDx™ Kit is regarded safe to use as one component in the determination of *HER2* protein overexpression in formalin-fixed, paraffin-embedded cancer tissue from patients with adenocarcinoma of the stomach, including gastroesophageal junction. In the BO18255 study *HER2* test results were missing for 16 patients (2.7%) with regards to FISH. The reasons for the unsuccessful *HER2* FISH were mainly due to not enough material available for testing. Another reason for unsuccessful *HER2* testing was that the samples did not pass quality control for various reasons.

From the data presented and discussed in this reproducibility study the most important conclusions drawn are:

- 1) Due to the heterogeneity in gastric cancer tissue of *HER2* gene amplification thorough scanning of the tissue is necessary to identify and enumerate the most amplified area and
- 2) For identification of all tissue cores in a biopsy sample it is recommended to inspect an H&E stained section.

In the effort to minimize the effect of heterogeneity and variability, additional recommendations for performance and interpretation of *HER2* FISH pharmDx™ 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 the reproducibility study results majority of variability can be attributed to the specimen due to tissue heterogeneity. When HER2:CEN-17 ratios are converted, agreement between observers and sites is high.

Nearly all patients enrolled in the trial had tumors which were gene amplified (FISH+) and only 16 patients were 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 16. 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. 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.

XV. REFERENCES

1. Bang YJ, et al. Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal

- junction cancer (ToGA): a phase 3, open-label, randomised controlled trial. *Lancet*. 2010; 376:687-697.
2. 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: 797-805.