INTERNATIONAL BREAST CANCER STUDY GROUP
IBCSG 45-13/BIG 4-13

PANACEA

A phase Ib/II trial evaluating the efficacy of MK-3475 (pembrolizumab) and trastuzumab in patients with trastuzumab-resistant, HER2-positive metastatic breast cancers

Statistical Analysis Plan

NCT02129556

August 11, 2015

Study Statistician: Anita Giobbie-Hurder

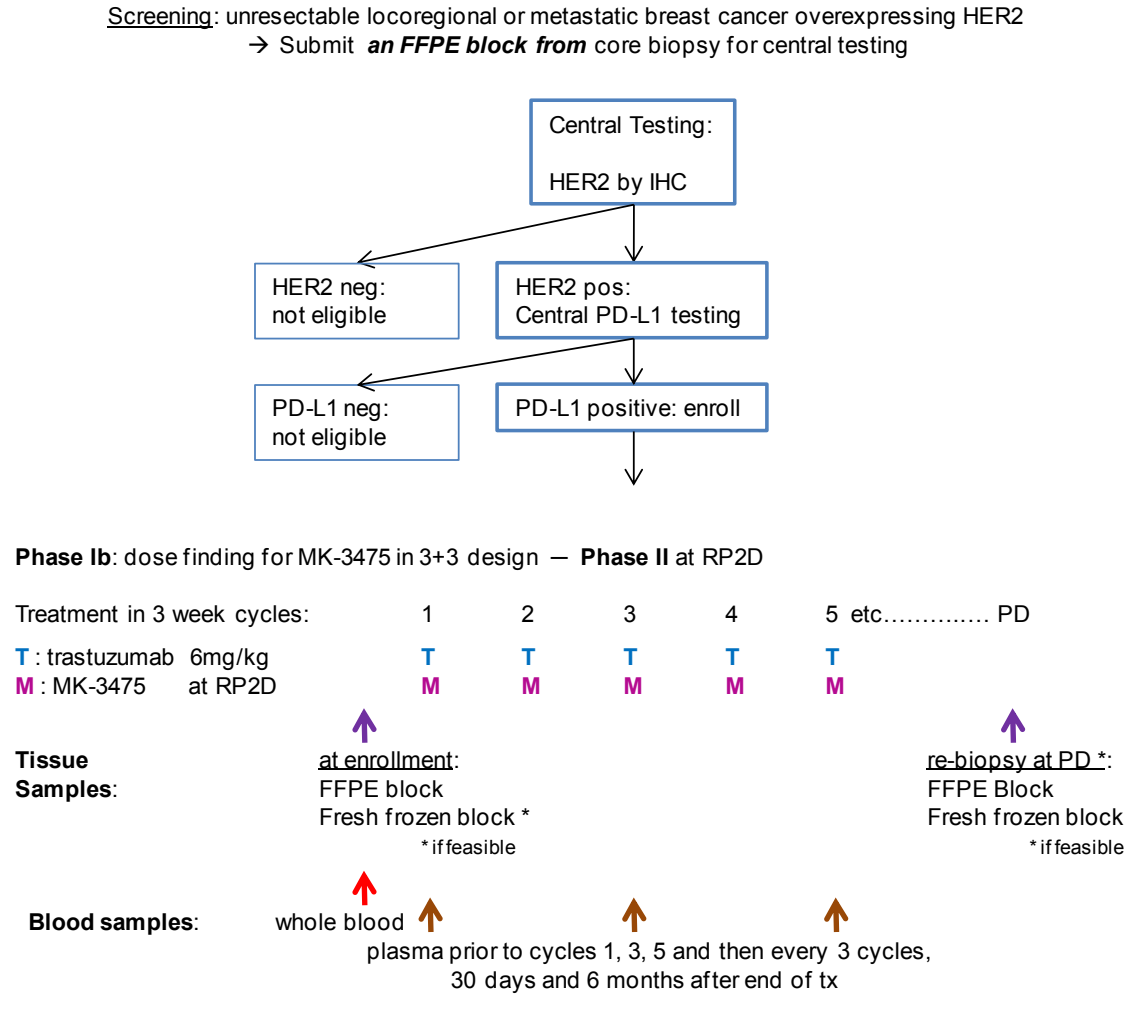
Table of Contents

1 Study Overview

1.1 Synopsis

This is a phase Ib/II clinical trial to evaluate the hypothesis that the combination of an immune re-activation approach and anti-HER2 therapy can reverse trastuzumab resistance and improve clinical outcomes in HER2-positive breast cancer. The primary objectives of this phase Ib/II study are to determine the recommended dose of the anti-PD-1 monoclonal antibody (mAb), MK-3475 (pembrolizumab), in combination with standard dose trastuzumab, and to evaluate the efficacy and safety profile of the drug combination in patients with PD-L1 expressing, HER2-positive unresectable loco-regional or metastatic breast cancer who have experienced progression during prior trastuzumab-based therapy. As part of Amendment #1 (QIII 2015), a second, parallel cohort of 15 patients with PD-L1 negative, HER2-positive, unresectable loco-regional or metastatic breast cancer will be enrolled in the phase II trial and evaluated for efficacy.

1.2 Schema



1.3 Statistical design assumptions and sample size considerations

It is expected that a maximum of 67 evaluable patients will be enrolled. The phase Ib trial will use a standard 3+3 dose-escalation design to determine the recommended phase-II dose (RP2D) of the anti-PD-1 mAb, MK-3475, in combination with standard dose trastuzumab using 3 possible dose levels (trastuzumab 6 mg/kg and MK-3475 at 2 mg/kg, 10 mg/kg, or a fall-back dose of 1 mg/kg). The RP2D, or a fixed dose of 200 mg MK-3475 will then be carried forward to the phase II trial. Each patient will receive combination trastuzumab and MK-3475 until disease progression, lack of tolerability, completion of 24 months of treatment with MK-3475, or until the patient declines further protocol treatment. In the phase II trial, 40 patients will be enrolled in the cohort of patients with PD-L1 expressing disease, and 15 patients in the cohort of PD-L1 negative disease.

Each patient's disease must be centrally assessed for expression of PD-L1 in a core biopsy taken from an unresectable loco-regional or metastatic lesion using Merck CLIA-certified IHC test in a designated lab. HER2 status of the tumor will also be centrally confirmed at the same time. A patient whose disease is both PD-L1 expressing and HER2-overexpressing (IHC 3+ or FISH/chromogenic in situ hybridization [CISH] positive) will be eligible for enrollment into the phase Ib or into the cohort of patients with PD-L1 expressing disease in the phase II. Patients with HER2-overexpressing and PD-L1 negative disease will be eligible for enrollment into the parallel phase II cohort.

Phase Ib: The sample size for the phase Ib trial will be between 6 and 12 enrolled patients and will depend on the number of dose cohorts needed to determine the RP2D. A patient in the phase Ib portion of the trial will be replaced if determination of dose-limiting toxicity (DLT; (defined in Protocol Section 10.2.1) cannot be adequately assessed because of rapid disease progression during the first cycle of therapy, or if treatment and/or follow-up is stopped during the first cycle of therapy for reasons other than toxicity. Patients who do not receive trastuzumab in the first cycle of therapy due to an infusion reaction to MK-3475 will also be replaced and will not be counted as having a DLT.

The following table summarizes the probability that the dose will be escalated (i.e., that 0 in 3 patients or 1 in 6 patients experiences DLT) given the true but unknown rate of DLT in cycle 1.

True, but Unknown, DLT Rate	Probability of Dose Escalation
10%	91%
20%	71%
30%	49%
40%	31%
50%	17%
60%	8%

Phase II:

PD-L1 Expressing: The phase II trial in the cohort of patients with PD-L1 expressing breast cancer employs Simon's optimal two-stage design, with a total target sample size of 40

evaluable patients. The null hypothesis of a true objective response rate (ORR) of 7% will be tested against a one-sided alternative objective response rate of 22%. Seventeen (17) evaluable patients will be enrolled in the first stage, and if one or fewer objective responses are observed in 17 patients, then enrollment will stop. If there are two or more objective responses, then an additional 23 patients will be enrolled, for a total of 40 evaluable patients. The null hypothesis will be rejected if a total of six or more objective responses are observed in 40 evaluable patients. This design yields a type I error rate of 0.05 (target type I error of 0.05) and power of 85% (target type II error of 0.15) when the true objective response rate is 22%. If the null hypothesis is true, the probability is 0.66 that enrollment will stop at the end of the first stage.

Enrollment will pause at the end of the first stage to allow for two, concurrent analyses of the data. One analysis will focus on the assessment of ORR using the criteria described above. The second analysis will be a detailed safety review, described in Protocol Section 11.5 and in Protocol Section 17.4. The enrollment pause will last until responses have been confirmed in the first 17 patients and safety has been evaluated carefully as described in Protocol Section 17.4.

PD-L1 Negative: The PD-L1 negative cohort is designed to provide evidence of response and inform subsequent drug development. A single-stage design with an enrollment of 15 patients will be used to compare a null response rate of 1% with a desirable response rate of 20%. The cohort size was selected to yield a very high probability of not missing a response signal if one exists. The decision rule is based on zero responses: the drug combination would not be considered worthy of further investigation if no patients respond. If the true response rate is 20%, then the probability that zero responses would be observed in 15 patients is 0.035. Therefore, there would be less than 4% chance of missing a true 20% response rate if the alternative hypothesis is true.

1.4 Screening and Enrollment

The total enrollment of this phase Ib/II trial will be between 6 and 67 evaluable patients. The minimum would occur if the combination of MK-3475 and trastuzumab proved too toxic at MK-3475 doses of 2 mg/kg and 1 mg/kg and the trial was stopped during the dose escalation phase. A maximum enrollment of 67 patients reflects two complete dose escalation levels of 6 patients each, and 55 patients (40 PD-L1-expressing and 15 PD-L1-negative) in the phase II. Under the assumption that approximately 50% of patients will be found ineligible upon screening, it is estimated that between 85 and 95 patients will be screened to obtain the maximum sample of 67 evaluable patients.

2 Overview of Data Analysis Software and Settings

Unless otherwise noted, all data manipulation and analyses will be performed using SAS. Categorical tables will be generated using PROC TABULATE with percentages calculated within column (COLPCTN). Fisher's exact testing will be based on two-sided exact probabilities produced using PROC FREQ with the EXACT option in the TABLE statement.

Tables of data measured on a continuous scale will be generated using PROC TABULATE and will include N, N missing, mean, standard deviation, minimum, median, and maximum for data summaries. Wilcoxon rank-sum testing will be conducted using PROC NPAR1WAY. Exact two-sided p-values will be used, whenever possible, which are generated using the EXACT WILCOXON statement.

Distributions of time-to-event outcomes will be calculated using Kaplan-Meier estimation from PROC LIFETEST. Confidence intervals will be estimated using the procedure default of log(-log(endpoint)). Numbers at risk will be shown (PLOTS=SURVIVAL (ATRISK)). The x-axis of these plots will be in months after start of therapy and the y-axis will be proportions.

3 Phase Ib

3.1 Analysis Population

All patients receiving one or more doses of MK-3475 and trastuzumab will be included in the assessment of DLT.

3.2 Primary Objective

To determine the recommended dose of MK-3475 in combination with standard dose trastuzumab.

3.3 Primary Endpoint

The incidence of dose-limiting toxicity of MK-3475 in combination with standard dose trastuzumab.

Dose escalation will employ a standard '3+3' approach, beginning in dose level 1, with rules for escalation and de-escalation described in the table, below. The RP2D is defined as the highest dose level at which less than 33% of patients (0 of three patients, or 0 or 1 of six patients) have experienced a DLT *in cycle 1*. Once dose escalation for MK-3475 reaches a dose of 10 mg/kg, no further escalation will occur. Intra-patient dose escalation is not permitted. Patients enrolled in the phase Ib part of the trial should be treated on the assigned dose level as long as tolerable, for up to 24 months of MK-3475.

Dose Escalation Rules

Number of patients with DLT at a given dose level	**Dose Escalation Rule**
0 out of 3	Proceed to the next dose level and enroll 3 patients
1 out of 3	Enroll and treat 3 additional patients at this dose level.
≥2 out of 3	Dose escalation will be stopped. The RP2D will be one dose below this dose level.
1 out of 6	Proceed to the next dose level.
≥2 out of 6	Dose escalation will be stopped. The RP2D will be one dose below this dose level.
If ≥2/3 or ≥2/6 patients at dose level 1 experience DLTs, dose level -1 will be used. If dose level -1 proves too toxic, the trial will stop.	

Data Sources

Purpose	CRF	Variable
Dose cohort	A2	ph1dl
No DLT	DLT	ndtox
DLT occurred	DLT	dltae1
Stopped during cycle 1 for reasons other than DLT	EoT	Fmno, msts
Other AEs for comparison	AE log	Assorted variables

DLT data are analyzed manually by the Medical Reviewer, Trial Biostatistician, and Data Management Center (DMC) team leaders. DLT results for each dose cohort are presented and discussed on scheduled conference calls. Reports regarding the observed toxicities, DLTs, and decisions about dose escalation are generated by the Biostatistician; distributed to the study PIs, DMC, and Coordinating Center; and stored in the study binder.

Presentation: Patient listing for all patients enrolled in phase Ib

Phase Ib Cohort	Patient ID	Enrolling Center	Screening Date	Enrollment Date	DLTs Documented	Other Cycle 1 Adverse Events

4 Phase II

4.1 Analysis Population

The phase II study population will be comprised of evaluable patients enrolled in the phase II trial. The PD-L1-expressing and PD-L1-negative cohorts will be evaluated separately.

4.2 Primary Objective

To evaluate the efficacy and safety profile of the drug combination in patients with HER2-positive, unresectable loco-regional or metastatic breast cancer who have experienced progression during prior trastuzumab-based therapy. Efficacy and safety will be assessed separately for the two PD-L1 expression cohorts.

4.3 Primary Endpoint

Objective response (confirmed CR or PR as best overall response) according to RECIST criteria (Version 1.1) as defined in protocol section 17.7.1.

At the time of each restaging, patients will be classified as achieving complete response (CR), partial response (PR), stable disease (SD), progressive disease (PD), or non-evaluable for response according to RECIST (Version 1.1) criteria. Objective response will be determined by

the best overall confirmed response designation recorded between the date of first dose of trial therapy and the date of objectively documented disease progression or cessation of trial therapy, whichever occurs first. For patients without documented progression or cessation of trial therapy, all available response designations will contribute to the objective response determination. The proportion of patients with an objective response will be presented with a two-sided 90% confidence interval calculated using the method of Atkinson and Brown, which allows for the two-stage design. The response rate in the PD-L1 negative cohort will be estimated and presented with a 90% exact, binomial confidence interval.

Data Sources

Purpose	**CRF**	**Variable Name**
Dose cohort	A2	ph1dl
	A2	ph2
Best Overall Response (BOR)	IR	?
Never started treatment	IR	?
No DLT	DLT	ndtox
DLT occurred	DLT	dltae1
Stopped during cycle 1 for reasons other than DLT	EoT	Fmno, msts

Analysis Steps:

1 – Identify patients who are enrolled in phase II;
2 – Separate into cohorts defined by PD-L1 expression;
3 – Document and remove any patients who never began treatment;
4 – Use BOR variable to classify response into responders vs. non-responders/unevaluable;
5 – Count responses or tabulate response rate, as appropriate for stage of analysis;
6 – Once trial complete use BCB "TWOCON" routine to calculate 90% two-stage confidence interval (Atkinson and Brown) in PD-L1-expressing cohort. Exact binomial 90% confidence interval for PD-L1 negative cohort will be estimated using the "CONFIN" procedure.

4.4 Secondary Endpoints and Definitions

There are several secondary endpoints for the trial, listed below. All efficacy analyses will be based on the phase II efficacy population, separate by PD-L1 cohorts; the safety analyses will be based on all patients receiving at least one dose of MK-3475.

1. *Safety and tolerability* according to NCI CTCAE version 4.0;
2. *Disease control (DC):* Best overall response of confirmed CR, PR, or SD lasting for 24 weeks or longer, measured from the start of trial treatment until first documentation of progressive disease;
3. *Duration of response (DoR):* Among patients with objective response (confirmed CR or PR as best overall response) as the interval between dates of first documentation of

objective response and first documentation of progressive disease. In the absence of documented progressive disease, follow-up will be censored at date of last disease assessment;

4. *Time to progression (TTP):* The interval between the dates of the start of trial treatment and first documentation of progressive disease. In the absence of documented progressive disease, follow-up will be censored at date of last disease assessment;

5. *Progression-free survival (PFS)*: Time from start of trial treatment until documented disease progression or death, whichever occurs first. Patients with new non-breast cancer malignancy must continue to be followed for progression of the original breast cancer. For patients without progression, follow-up will be censored at the date of last disease assessment without progression, unless death occurs within 12 weeks following the date last known progression-free, in which case the death will be counted as a PFS event. Patients who discontinue or initiate non-protocol treatment prior to documented disease progression will be followed for disease progression;

6. *Overall Survival (OS)*: Time from start of trial treatment to death from any cause. For patients who are lost to follow-up or who have no documentation of death at the time of final analysis, follow-up will be censored at the date of last assessment of vital status.

4.4.1 Safety and Tolerability

4.4.1.1 Early Safety Review (may be modified slightly pending Amendment)

An early safety review will be concurrent with the first efficacy review of ORR that is specified by the Simon two-stage design in the PD-L1 expressing cohort. Details regarding this early safety review are presented in detail in Protocol Sections 11.5 and 17.4. Data from the initial 17 patients receiving the combination of MK-3475 and trastuzumab in the phase II study will be evaluated for safety in the first 4 cycles of therapy. After this initial safety assessment, toxicity and adverse event data will be evaluated as part of the semi-annual trial review by the DSMC.

As part of the early safety monitoring, we will focus particular attention on cardiac events (CE) of grade 3 or higher (Protocol Section 11.5). The incidence of the following cardiac events will be analyzed:

- Cardiac death
- Heart failure manifested by dyspnea with normal activity or at rest
- Decline in LVEF of more than 10 percentage points from baseline to a value less than 50 percent.

If three or more of the first 17 patients experience CE, then the trial enrollment may be suspended. If the true probability of cardiac toxicity is 18% or higher, the probability is greater than 60% that three or more patients will be observed with CE at the time of the early safety monitoring.

All reported AEs will also be summarized as part of this safety review. For a sample of 17 patients, there is high probability of observing at least one event if a toxicity has an incidence of at least 9%. If the true incidence of an unexpected or severe toxicity is 9% or greater, the

probability is at least 0.80 that one or more patients out of 17 will experience the toxicity during the early safety monitoring period. If the true incidence of unexpected or severe toxicity is 3% or less, the probability is 0.40 or less that at least one patient of 17 will experience the toxicity during the early safety monitoring period.

Analysis Steps for Early Safety Review:

1 – Identify patients in the appropriate phase II cohort and that are evaluable, having received at least one dose of study therapy;
2 – Restrict and examine AE and SAE/ECI CRF data for first four cycles of therapy

Cardiac Event Assessment:

3 – AE CRF CTCAE v4 codes (others may be added):
Cardiac death – CA108
Heart failure – CA111, CA112, CA125, CA126
LVEF dysfunction – CA113
Other cardiac grade 3 or higher – CTCAE code beginning with "CA"
4 – Examine SAE/ECI data for first four cycles of therapy, specifically for cardiac events
Reason for seriousness = "Constitutes Important Event", or
"Event is an ECI"

All Other Reported Adverse Events:

5 – Determine highest grade of each CTCAE code reported during the first four cycles of therapy
6 – Compare with baseline AEs (BAE CRF) and remove those that have continued since baseline without worsening grade
7 – Examine SAE/ECI data and cross-check with events reported on AE CRF. Remove duplicates
8 – Summarize according to worst grade reported for each patient

Presentation:

1 – *Cardiac Event Assessment*

a. Table of cardiac events (by CTCAE term) according to worst grade reported and relationship to therapy
b. Identify if ECI or SAE
c. Provide patient listings separately by worst grade for grades 3, 4, and 5 with action taken
d. Calculate number of patients with one or more grade 3 or higher cardiac events and present with 90% exact confidence interval.

2 – *All Other Reported Adverse Events*

a. Table of CTCAE terms according to worst grade reported and relationship to therapy

b. Identify if classified as SAE or ECI
c. Provide patient listings by worst grade for each event of grade 3, 4, or 5 with action taken
d. Calculate number of patients with one or more grade 3 or higher non-cardiac events and present with 90% exact confidence interval.

4.4.1.2 Subsequent Reviews of Safety

Toxicity and adverse event data will be evaluated semi-annually for review by the DSMC. Any patient who received one or more doses of the study therapy will be included in the evaluation of safety. Processing of data and methods of summary will be comparable to those described for the Early Safety Review with adverse events and toxicities identified via CTCAE code and summarized according to the worst grade reported for each event type per patient. Separate tables will be presented according to cohort (Phase Ib, Phase II PD-L1 expressing, Phase II PD-L1 negative), and, where applicable, by dose. If deemed appropriate by PIs and Medical Reviewer, safety data for the PD-L1 expressing and negative cohorts may be combined and presented together for the Phase II trial.

4.4.2 Disease Control Rate

The rate of disease control is the proportion of patients in the Phase II trial with a best overall confirmed response of CR or PR, or SD lasting for at least 24 weeks. All times are measured from the first dose of study therapy. Duration of SD will be determined by the first documentation of confirmed progressive disease. Patients who are unevaluable for response are included in the analysis and are counted in the rate denominator. Disease control rates will be summarized by PD-L1 cohort.

Data Sources

Purpose	CRF	Variable Name
Dose cohort	A2	ph1dl
Best Overall Response (BOR)	IR	?
Never started treatment	IR	?
Stable Disease ≥ 24 wks	IR	?

Analysis Steps:

1 – Identify patients in Phase II cohort who had at least one dose of study therapy

2 – Verify BOR against Tumor Evaluation Forms (TEV) to confirm response over time and duration of SD

3 – Use BOR and SD variables to classify patients into disease control vs. no disease control

4 – Calculate disease control rate and present with exact 90% binomial confidence interval

4.4.3 Time-To-Event Endpoints

There are four secondary, time-to-event endpoints: duration of response, time to progression, progression-free survival, and overall survival. The timing of the last three begins at the first dose of study therapy. Duration of response is based on a subset of patients and is timed from the first date of confirmed CR/PR. All times will be expressed in months. Summaries will be presented by PD-L1 cohort.

4.4.3.1 Duration of Response

Data Sources

Purpose	**CRF**	**Variable Name**
Dose cohort	A2	ph1dl
Best Overall Response (BOR)	IR	?
Never started treatment	IR	?
Stable Disease ≥ 24 wks	IR	?
Date CR/PR first observed	IR	?
Criteria for PD?	IR	?
Date PD observed	IR	?
Date last adequate disease assessment	IR	?

Calculation Steps

1 – Identify patients in Phase II cohort who had at least one dose of study therapy
2 – Restrict sample to patients with CR or PR as best response to therapy
3 - Patients with PR followed by CR will use the date of PR as the reference date
4 – For patients with subsequent PD,
    DoR = (date progression – date CR/PR)/30.4375.
    These patients will be classified as having an event (censor = 1)
5 – For patients without subsequent PD,
    DoR = (date last adequate assessment – date CR/PR)/30.4375.
    The follow-up of these patients will be censored (censor = 0).

4.4.3.2 Time to Progression

Data Sources

Purpose	CRF	Variable Name
Dose cohort	A2	ph1dl
Treatment start date	IR	?
Never started treatment	IR	?
Criteria for PD?	IR	?
Date PD observed	IR	?
Date last adequate disease assessment	IR	?

Calculation Steps

1 – Identify patients in Phase II cohort who had at least one dose of study therapy

3 – For patients with subsequent PD,

TTP = (date progression – date treatment start)/30.4375.

These patients will be classified as having an event (censor = 1)

4 – For patients without subsequent PD,

TTP = (date last adequate assessment – date treatment start)/30.4375.

The follow-up of these patients will be censored (censor = 0).

4.4.3.3 Progression-Free Survival

Data Sources

Purpose	CRF	Variable Name
Dose cohort	A2	ph1dl
Treatment start date	IR	?
Never started treatment	IR	?
Criteria for PD?	IR	?
Date PD observed	IR	?
Date last adequate disease assessment	IR	?
Date death	IR	?

Calculation Steps

1 – Identify patients in Phase II cohort who had at least one dose of study therapy

2 – For patients with subsequent PD,

PFS = (date progression – date treatment start)/30.4375.

These patients will be classified as having an event (censor = 1)

3 – For patients without subsequent PD and last-known alive,

PFS = (date last adequate assessment – date treatment start)/30.4375.

The follow-up of these patients will be censored (censor = 0).

4 – For patients without subsequent PD and who died:

a. If (date death – date last assessment)/7 ≤ 12,
   PFS = (date death – date treatment start)/30.4375.
   These patients will be classified as having an event (censor = 1).
b. If (date death – date last assessment)/7 > 12,
   PFS = (date last assessment – date treatment start)/30.4375.
   The follow-up of these patients will be censored (censor = 0).

4.4.3.4 Overall Survival

Data Sources

Purpose	**CRF**	**Variable Name**
Dose cohort	A2	ph1dl
Treatment start date	IR	?
Never started treatment	IR	?
Date death	IR	?
Date last adequate disease assessment	IR	?
Date last contact with patient	E	source

Calculation Steps

1 – Identify patients in Phase II cohort who had at least one dose of study therapy
2 – For patients alive at last follow-up,
   OS = (max(date last adequate assessment, source) – date treatment start)/30.4375
   The follow-up of these patients will be censored (censor = 0).
3 – For patients who died,
   OS = (date death – date treatment start)/30.4375
   These patients will be classified as having an event (censor = 1).

Presentation of Time-To-Event Results

The distributions of duration of response, time to progression, progression-free survival, and overall survival will each be summarized using the product-limit method of Kaplan-Meier. Median times for each endpoint will be presented with two-sided 90% confidence intervals estimated using log(-log(endpoint)) methodology. Kaplan-Meier estimates of TTP and PFS at 6 or 12 months after treatment initiation will also be presented with two-sided 90% confidence intervals.

5 Additional Presentations and Analyses

5.1 Accrual

Graphics:

1. Number of patients screened and enrolled by month
2. Cumulative number of patients screened and enrolled by month
3. Numbers of patients screened and enrolled by center

Tables: Overall and by Cohort

1. Reasons for ineligibility/non-participation

5.2 Study Sample

Table: Overall and by Cohort

1. Patient status including attrition before measurement of primary outcome
2. Reasons for attrition
3. Numbers of deaths, causes of death
4. Changes in consent
5. Number of cycles of study therapy (MK-3475, trastuzumab)
6. Reasons for stopping therapy
7. Receipt of non-protocol, systemic, anti-tumor treatment prior to progression

5.3 Patient Demographics and Baseline Disease and Prior Treatment

Tables: Overall and by Cohort

1. Age at enrollment
2. Race
3. ECOG performance status
4. Menopausal status
5. Time (months) between enrollment and
    a. Diagnosis of advanced breast cancer
    b. Diagnosis of most-recent disease progression
    c. Diagnosis of primary tumor
6. Trastuzumab and anti-HER2 therapies prior to enrollment
7. Prior endocrine therapy for metastatic disease
8. Prior radiotherapy for metastatic disease
9. Prior chemotherapy
10. Histologic type (ductal, lobular, not determinable)
11. Baseline tumor burden (sum diameters of all target lesions)
12. ER status (percent expression), PgR status (percent), Ki-67 (percent)
13. Tumor-infiltrating lymphocytes (percent)

14. If detail available, percent PD-L1 espression

5.4 Patterns of Response

Graphics:

1. Spider (spaghetti) plot of tumor responses over time based on the sum of diameters of target lesions. Will include annotations of new tumors and pseudo-progressions. Plots will be separated by cohort.
2. Waterfall plot of best overall tumor response. Cohorts will be color coded.

5.5 Correlatives

Proposed correlative data analyses are based on patients enrolled in the PD-L1-expressing cohort of the Phase II trial. Comparable analyses may be performed in the cohort of patients with PD-L1 negative disease. If deemed appropriate by the PIs and Medical Reviewer, the two samples may be combined in additional correlative analyses.

5.5.1 Tumor-Infiltrating Lymphocytes (TILs)

To examine responses according to pre-treatment levels of TILs, the population will be divided retrospectively according to objective response (CR/PR) or non-response. Pre-treatment percentages of stromal infiltrating lymphocytes will be summarized descriptively for the two response groups and compared using Wilcoxon rank-sum tests. Based on a sample of 40 patients, if there are 6 responses and 34 non-responses, a Wilcoxon rank-sum test with a two-sided, 10% type I error will have 85% power to detect a difference in baseline lymphocyte percentage that is 1.2 times the common standard deviation.

Visualization of the relationship between baseline TILs and the distributions of TTP or PFS will be based on Kaplan-Meier estimates stratified by lymphocyte-predominant breast cancer (LPBC) phenotype or median of the distribution of intratumoral or stromal percentages. LPBC phenotype will be defined as 50% infiltration of either stromal or intratumoral lymphocytic infiltration. Medians of the time-to-event endpoints will be shown with two-sided 90% confidence intervals; the distributions of TTP or PFS will be compared across TIL strata using the log-rank test.

Changes in TILs between baseline and progression/treatment discontinuation will be calculated (post-pre) for each patient and summarized descriptively.

5.5.2 Estrogen Receptor Expression

Pre-treatment ER expression will be dichotomized as present (≥1% expression) or absent (<1%). The proportions of patients with objective response (CR/PR) in each ER subgroup will be summarized with two-sided 90% exact, binomial confidence intervals. The distributions of TTP

or PFS will be summarized using Kaplan-Meier estimates and compared across the ER strata using the log-rank test.

5.5.3 Outcome According to Quantified HER2 Level

The study population will be divided retrospectively according to objective response or non-response. Pre-treatment FISH ratios or HER2 copy numbers will be summarized descriptively for the two response groups and compared using Wilcoxon rank-sum tests. If there are 6 responses and 34 non-responses, a Wilcoxon rank-sum test with a two-sided, 10% type I error will have 85% power to detect a difference in measure that is 1.2 times the common standard deviation.

Visualization of the relationship between FISH ratio or HER2 copy number and the distributions of TTP or PFS will employ Kaplan-Meier estimates. FISH ratio or HER2 copy number data will each be divided into high/low groups at the medians of the respective distributions. The distributions of TTP or PFS will be compared across FISH ratio or HER2 copy number strata using the log-rank test; medians of the time-to-event endpoints will be shown with two-sided 90% confidence intervals.