Adaptive seamless phase II/III study in gastric cancer (orphan condition)

25-June-2015
The GATSBY seamless phase II/III study been initiated and guided by Dr. Michael Budde from the Biostatistics Department at F. Hoffmann-La Roche.

Maximo Carreras (F. Hoffmann-La Roche) been the statistician involved with the study planning (including extensive data simulations) to write his PhD in collaboration with Prof. Dr. Werner Brannath from the University of Bremen.
Seamless Adaptive Phase II/III

Traditional Phase II + Phase III trials

- Control
- Group 1
- Group 2

Phase II

Data analysis

Planning Phase III

End of Phase III

Seamless Adaptive Phase II/III trials

- Control
- Group 1
- Group 2

Phase II

Development Time

End of Phase III
• Patients with locally advanced or metastatic HER2-positive gastric cancer, second-line treatment
• Patients randomized 2:2:1 ratio to Trastuzumab emtansine 3.6 mg/kg (q3w), Trastuzumab emtansine 2.4 mg/kg (qw) or control
• Primary endpoint: Overall Survival
• Median time to OS assumed 6 months in control and 9 months in both treatment groups (HR=0.67)
GATSBY Study

Population:
2L HER2+ AGC
Prior chemo ± prior HER2 tx
PS 0–1
Total n=412 (Stage 1 + 2)

Randomization:
Stage 1: 3 Arm; 2:2:1 ratio
Stage 2: 2 Arm; 2:1 ratio

* Regimen selection based on PK, efficacy and safety
**Stage 1 (Stage 2) patients consist of all patients recruited before (after) the regimen selection decision
• Clinical cut-off for dose selection by an iDMC, after 100 patients across all three arms treated for a minimum of 12 weeks

• Accrual continues into all 3 treatment arms until dose selection. All patients randomized before the selection considered “stage 1 patients”

• After the selection remaining patients recruited into the selected T-DM1 arm and the control in a 2:1 ratio

• Case of delay for group selection -> more patients in stage 1 and less in stage 2
Recruitment ongoing after clinical cut-off for dose selection and before iDMC decision (~2 months in GATSBY)

Control (Taxane)  
- n=30  
- n=6  
- n=77  
- Total n=113

Selected Arm (2.4 mg/kg T-DM1 qw)  
- n=60  
- n=12  
- n=154  
- Total n=226

Non-selected Arm (3.6 mg/kg T-DM1 q3w)  
- n=60  
- n=12  
- Total n=72
GATSBY Study Timelines

- **FPI Stage 1**
  - 2012 Q3
  - ~9 months
  - 100 patients randomized + 12 weeks follow-up

- **Stage 2 patient enrollment started**
  - 2013 Q3
  - ~2 Months*

- **Regimen selection decision for stage 2**
  - 2013 Q4
  - ~5 months

- **Clinical cut-off Futility Analysis (After ~ 68 events)**
  - 2014 Q1
  - x months

- **Clinical cut-off Main efficacy analysis**
  - 2015

* *Recruitment continues during this period*
Dose Selection by iDMC

- Guidance needed within the iDMC charter to be elaborated enough so that iDMC can make appropriate recommendations (decision to some extent with iDMC and not the sponsor)
- In GATSBY all available safety, PK, and efficacy data were provided to be used by the iDMC to make their recommendation on which trastuzumab emtansine regimen should be selected
- For sample size calculations a simulated regimen selection analysis was performed based on cut-off values for specific safety, PK and efficacy parameters (considered as good surrogates for the totality of the data)
Dose Selection by iDMC

1. Mean cycle 1 trastuzumab emtansine AUC for 3.6 mg/kg q3w regimen > 195 day.ug/mL

   **YES**

   Efficacy of 3.6 mg/kg q3w regimen inferior to that of 2.4 mg/kg qw regimen (*)

   **NO**

   **YES**

   Safety profile of 2.4 mg/kg qw regimen is appropriate

   **NO**

   **YES**

   Select 2.4 mg/kg qw regimen (***)

   **NO**

   **YES**

   Select 3.6 mg/kg q3w regimen (**)(***)

   (*) Collective expertise and judgment of IDMC members should be used.
   (**) Provided that safety profile of 3.6 mg/kg q3w regimen is appropriate.
   (*** Efficacy should always be considered in the final regimen selection.
Operational Aspects

- Sites selected with high recruitment at start of stage 1 (need of sufficient data within short time)
- SMT to provide clean data within short timelines (#patients recruited after clinical cut-off but before iDMC decision to be as small as possible)
- Change in randomization to be set-up and performed right after dose selection
- After dose selection, all patients to continue treatment/follow-up as before (including non-selected arm)
- Needed understanding by investigators and health authorities about staying blinded to the interim data the dose decision is based on
FDA draft Guidance for Industry (2010)

• FDA definition of adaptive design clinical study: “a study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study”

• Chief concerns “are control of the study-wide Type I error rate, minimization of the impact of any adaptation-associated statistical or operational bias on the estimates of treatment effects, and the interpretability of trial results”

• “less well understood”* design/methods “are primarily intended for circumstances where the primary study objective(s) cannot be achieved by other study designs”

* Any design involving interim review of unblinded efficacy data, beside group sequential analyses
Stage 1 consists of patients recruited before dose selection (two arms and control group)

For final analysis, after pre-specified number events, in a first step p-value calculated for stage 1 patients

Although interest in testing selected arm to control – as per Health Authorities main concern on type I error, here we need to account for multiplicity!

Due to available and valid follow-up of patients in the non-selected arm – this data can be used for inclusion to the final testing procedure
Patients in Non-selected Arm

In GATSBY all stage 1 patients continue to be treated according to therapy they were randomized to receive and followed up by pre-specified period of time – including patients in non-selected group after selection (possible due to orphan condition and OS being primary endpoint).

Control

n = xx

Selected arm

n = xx

n = xx

Non-selected arm

n = xx

End of Stage 1: iDMC Group Selection

End of Stage 2: Final Analysis

Treatment and follow-up of patients in non-selected group – shall only planned to be used if not be affected after selection!
Statistical Method Controls Type I Error
Stage 1 Part (Hochberg Correction)

Selected arm p-value > Non-selected arm p-value?

No

2*selected arm p-value > Non-selected arm p-value?

Yes

Stage 1 p-value = Max(2*non-selected arm p-value, selected arm p-value)

No

Yes

Stage 1 p-value = Non-selected arm p-value

No

Stage 1 p-value = 2*selected arm p-value
Statistical Method Controls Type I Error

P-value derivation: Stage 1 and 2

- Stage 2 consists of patients recruited after dose selection (only selected arm and control group)
- P-value for stage 2 patients is calculated and combined with stage 1 p-value - similar to combination of p-values in meta analysis
- Basic Assumption: Test statistics of stage 1 and stage 2 are independent (p-values uniformly distributed under null hypothesis)
- Regards the basic assumption (Jenkins et al, 2011): It is not permissible for stage 1 follow-up to be affected by the stage 2 design
Plan is for approximately identical end of stage 1 and 2, but depending on actual event rates cut-off dates may be different.

Such difference means: For the primary analysis we need to disregard events if higher (e.g. >5%) from the actually used data cut-off compared to the study plan.

**Stage 1 may reach cut-off first:**

- Control: \( n = xx \)
- Selected Arm: \( n = xx \)
- Non-selected: \( n = xx \)

End of Stage 1: iDMC Group Selection

End of Stage 2: Final Analysis

**.. or stage 2 may reach cut-off first:**

- Control: \( n = xx \)
- Selected Arm: \( n = xx \)
- Non-selected: \( n = xx \)

End of Stage 1: iDMC Group Selection

End of Stage 2: Final Analysis
Stage 1 and Stage 2 p-values to be combined for final test

Used Inverse normal combination test (as described in Jenkins et al (2011))

\[ P_{(S,fn)} = 1 - \Phi(Z_{fn}), \quad \text{with} \quad Z_{fn} = w_1 \left( 1 - P_{S1} \right) + w_2 \left( 1 - P_{S2} \right) \]

\[ w_1 = \sqrt{ \frac{e_1}{e_1 + e_2} } \quad \text{and} \quad w_2 = \sqrt{ \frac{e_2}{e_1 + e_2} } \]

with \( e_1 \) and \( e_2 \) expected events from stage 1 and 2, respective

\( w_1 \) and \( w_2 \) need to be fixed before start of study (fixed per statistical analysis plan, based on the results of the simulation study: \( e_1 = 166, \quad e_2 = 134 \) with \( n_1=200, \quad n_2=212 \))
Statistical Method Controls Type I Error
Final Testing Procedure

Any p-value* combinations within area under the curve can be rejected

*Use of one-sided p-values to exclude use of reverse/detrimental effects
HA Further Concern: Bias in Estimates

- Investigation of bias by means of simulations encouraged by CHMP and bias in worst-case analysis shall be performed to provide upper bound
- Worst case assumption: Same treatment effect in both regimens and selecting study arm with larger effect size with already knowing the «future» final outcome (due to review of surrogate endpoints) – unrealistic perfect surrogate, but can be simulated
# Bias in Estimates

Worst case assumption bias (regards selected arm) at the planned final analysis cut-off with simulated median time to event for both treatment arms 9 months and control 6 months

<table>
<thead>
<tr>
<th>Hazard Ratio</th>
<th>Median time to event, selected arm (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Bias</td>
<td>SE</td>
</tr>
<tr>
<td>-.014</td>
<td>.00012</td>
</tr>
</tbody>
</table>

Based on 500,000 study simulations

Investigation of bias in estimates resulting in only small deviation. An observed HR of 0.67 may be corrected to HR of 0.684, and median time to OS subtracting one week.
The Study Estimates are derived similar to a usual phase III study – using all data within selected arm and control.

Versus for the statistical testing a derivation is needed based on stage 1 and stage 2 patients separately and with including the non-selected arm.
Final Reporting
How to Explain the Selection?

At point in time of regimen selection only data until past cut-off been used (information only available to iDMC until final reporting)

Versus at point in time of final reporting additional data is available
Reporting in GATSBY

Phase III

Phase II

Stage 1

Stage 2

Control

Selected Arm

Non-selected Arm

n=30

n=6

n=60

n=12

n=60

n=12

Total n=113

Total n=226

Total n=72

n=77

n=154

iDMC CCoD

iDMC selection

Final Analysis
• Phase II: Data from Stage 1 patients up to clinical cut-off used for IDMC regimen selection
• Phase III: Combined Stage 1 and Stage 2 data for the confirmatory portion of the study up to final cut-off. Control arm and the selected arm presented in two-column, side-by-side tables. The non-selected arm presented separately.
• OS endpoint analysis incorporates p-values from both study stages. OS and demography from all patients enrolled in Stage 1 will be shown in three-column, side-by-side fashion.
Interim Analysis

- An interim analysis to allow stopping for futility is possible
- Interim efficacy analysis problematic, due to the complex final testing procedure

1. Scenario stopping: Due to 2 study stages the planned final testing not applied at interim $\rightarrow$ treatment effect to be very overwhelming to account for this lack in control of study-wide type I error rate (GATSBY: HR < 0.357 in 68 events)

2. Scenario not stopping: The used alpha at interim need to be split (instead use of nominal alpha) $\rightarrow$ possibly higher loss of alpha (GATSBY: alpha spent=0.00003 with expected 68 events)
# GATSBY Design Pros and Cons

<table>
<thead>
<tr>
<th>Pros</th>
<th>Cons</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time/cost saving (Phase II/III in one study)</td>
<td>Operationally higher challenge and complex simulations for design/power evaluations</td>
</tr>
<tr>
<td>Using all data for dose selection with alpha fully controlled</td>
<td>Non-selected group to continue with the pre-planned treatment</td>
</tr>
<tr>
<td>Only small bias in estimates</td>
<td>Bias needs to be investigated/reported</td>
</tr>
<tr>
<td>Dose selection by iDMC*: Roche independent expertise</td>
<td>- iDMC: Challenging «black box» with risk</td>
</tr>
<tr>
<td></td>
<td>- Great time gap between decision and reporting</td>
</tr>
<tr>
<td>Phase II/III in only one reporting</td>
<td>Non-selected arm vs control and/or by stage reporting (selected arm) may get HA request</td>
</tr>
<tr>
<td>Interim stop for futility possible</td>
<td>Interim stop for efficacy not straightforward</td>
</tr>
</tbody>
</table>

* Essential for adaptive study design - to keep the sponsor blinded
Summary

• GATSBY is first Roche oncology study to apply this type of design
• It is an experimental study design («less well understood» design by FDA), but it is cost and time saving with type one error fully controlled
• The design used in GATSBY is not generalizable: Specifically the early group selection by an iDMC and continued treatment in the non-selected arm after the selection may not always be possible (primary endpoint in GATSBY is overall survival in patients with no promising other treatment options)
References


Back-up
Simulation: Dose Exclusion Criteria

Safety: Drop a T-DM1 arm if
- Treatment related mortality (TRM) rate ≥ 5% than control arm
- TRM > 10% regardless of comparison to control arm

PK: Drop a T-DM1 arm if
- Cycle 1 AUC is < 65% of historical control observed for the same schedule in mBC

Efficacy: Drop a T-DM1 arm if
- HR for OS > 1.5 vs. control arm

Both T-DM1 arms may be dropped if they both meet any of the above criteria
Comparison is done if neither T-DM1 arm is dropped based on pre-defined criteria (previous slide):

- **PK**: If one dose shows 50% or higher dose-intensity adjusted cycle 1 AUC, select that dose. Otherwise efficacy and safety should be considered

- **Efficacy**: Dose with lower HR for OS (vs. control) should be selected

- **Safety**: Dose with lower TRM rate and treatment-related AEs should be selected. In addition, dose intensity may be considered in overall treatment safety assessment. Dose schedule with higher dose intensity is preferred

Note: Dose-intensity adjustment and treatment related AEs not used in simulations
Simulating Bias in Worst-case Analysis

1) Events simulated of all arms using protocol assumptions with same effects in the experimental arms and censored at pre-defined end of stage 2

2) Use of simulated stage 1 patients randomized before pre-defined clinical cut-off of the group selection: Comparing the hazard ratio (HR) for group 1 vs control to group 2 vs control to select group with smaller HR

3) For group selected: Calculate the final HR using simulated data of all stage 1 and 2 patients combined and compare to protocol assumed HR
The family-wise error rate (FWER) can be inflated when testing primary and secondary endpoints. Specifically testing an endpoint used for regimen selection and scenario of only one arm with OS effect ($1 + (2$ need to be controlled:

1) Probability arm with no OS effect get selected (worst assumption: surrogate independent to primary) and primary testing wrongly rejected (may approach “cost” of full alpha due to inclusion of the possibly very small p-value of the non-selected arm with effect)

2) Probability treatment arm with effect get selected and secondary endpoint wrongly rejected

An adaption of the more conservative Dunnett testing procedure$^1$ as described in König et al (2008) prevents «cost» of full alpha from scenario above point 1 (if used for the primary analysis) due to setting the $z$-statistic for the deselected treatment to infinity (using this test for the final analysis -> the non-selected arm does not need to be followed up)

1) Requires overall number of events to be pre-fixed (can not be changed at interim)
Futility Analysis Operational Aspect

- Futility analysis to be performed after approximately 68 events within selected arm and control
- Sponsor shall be blinded to calculations by arm, therefore only number of events over all 3 arms can be investigated
- Theoretical assumptions from study start need to be used for timeline projections (assumption on identical event rates between experimental arms)
- For GATSBY the projection for futulity analysis timeline been presented to the iDMC during the regular safety reviews