Subgroup selection for biomarker
An overview

Hans Ulrich Burger
Roche MCO group

Leiden, June 24, 2016

Overview

- Overview
  - Impact of biomarker on development programs
  - Current frequent problems
- Companion diagnostics development
  - Biomarker cut-off selection
- Frequent requirements and solutions
  - Companion diagnostics
  - Enrichment decisions
  - Role of biology
  - Role of surrogate endpoints
- Summary

My context for today

- As the situation is complex my goal for this talk was rather
  - To highlight complexity
  - To provide overview on what is done today
  - To highlight areas where more is required
- The objective is however not
  - To present any new methods
  - To present examples in depth

Overview
Why enrichment?

- Three scenarios when developing a new molecule today
  - Enriched patient population for sure (for example trastuzumab)
  - Potentially enriched patient population defined by a few biomarkers (BM) (for example bevacizumab)
  - Enriched population unlikely with/out potential exploratory biomarker analyses (for example rituximab)
- Why enriched patient population?
  - Larger effect leading to larger benefit for patients
  - Larger benefit leads to higher value for patients and societies
  - De-risking strategy in case overall population does not work
- This is not only oncology!

Overview
Why enrichment?

- Development relatively straight forward in first and third case:
  - In first case clinical development program nearly unchanged, but additional complexity due to parallel companion diagnostic program and perhaps cutoff selection needed for BM
  - In third case clinical development program unchanged, just exploratory biomarker analyses potentially added (if at all)
- But how to design a development program in second case still unclear without good solution. This needs further discussion
- Can we ignore BM?
  - No, because of patient needs
  - Predictive BM increases development risks by higher variability between studies

Overview
List of current problems

- Companion diagnostics
  - Assay development: Develop assay with the right properties
  - Consistency between different versions of assays used?
  - Assay development is a challenge as development usually in parallel to phase 2 and 3

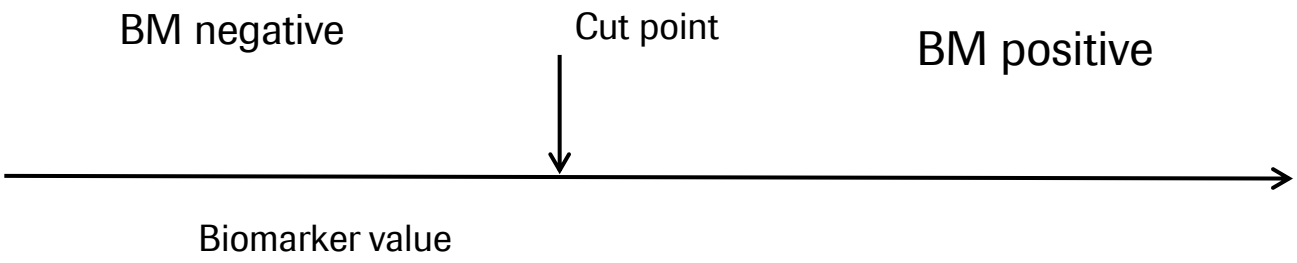
- Decision making on enrichment
  - Is there an enriched population?
  - If yes driven by which BM?
  - Differentiate between prognostic and predictive BM
  - Cut-off selection for the BM?
    - Can predictive cut-offs be defined on prognostic value only?
  - Role of surrogate endpoints?
  - What is the right time point for BM decision making, phase 2 or 3?

Overview
List of current problems

- Multiplicity
  - Huge problem: Many BMs looked at in one or two studies leading to high rate of false positive
- Regulatory requirements
  - How much data on BM negative patients required?
  - Qualitative versus Quantitative interactions
  - How much data is required for enrichment decision ?
- Role of biology sometimes strong
  Note: Role of biomarker varies across diseases including oncology
  - Simple example:
    Test for bacterial infection to guide the use of antibiotics

Compagion diagnostic development

- Cut point determination looks like we have a one dimensional scale in which we need to define the right cut for subgroup definition:



- Reality however often more complex
  - Frequently, BM is a receptor for a certain kind on tumor cells:
    - BM does not mean that all cells do not have the receptor. The number of receptors at a cell and the number of cells with receptors could vary and all this may not be stable over time => Multidimensional!

Companion diagnostics development

- Example: HER 2 biomarker stable over time but this seems rather the exception than the rule
  - Generally, not given that a BM negative patient this time will be negative also next time. Consequence?
- Frequently, BM and new therapy developed the same time and no validated assay yet ready
  - Development of assay usually first but its validation runs in parallel
  - Some of the clinical data based on older assay
  - Additional uncertainty due to variability in assays
- Validation done usually in phase III without major impact on critical study design features

Companion diagnostics development
Biomarker cut off selection

- In the simple case: Biomarker cutoff selection often not a major issue in terms of design but size
  - BM distribution homogenous
  - BM continuous on a one dimensional scale

  => Main problem then only the limited information
- Cut off selection frequently based on categorization into 0, 1+, 2+ and 3+ to overcome biological assay challenges especially when biomarker effect not only on one scale
  - Categorization however often difficult and multidimensional
  - Categorization often not consistent and stable in time
  - Categorization however useful as it drives for manageable decisions

Companion diagnostics development
Biomarker cut off selection

- Options available:
  - Minimize p-value for interaction test between BM and treatment
  - Maximize interaction statistic between treatment and BM
  - Maximize the difference in treatment effect in BM+ versus BM-
  - Look at BM- group only and define cutoff under the condition of no treatment effect in BM- subgroup
  - ….
- Cut off selection should be primarily done to exclude a clinically relevant effect in BM- population
- Further complicating factor:
  Cut off definition also driven by prevalence of BM+ subgroup and business case

Companion diagnostics development
Biomarker cut off selection

- Assume BM also prognostic, then the prognostic value could be used to define cutoff values (Faraggi&Simon, Stat med, 1996, 2203-2213)
  - Split study population by cut off value x in patients with BM > x and BM ≤ x and test for difference
  - Take cut off value x_0 which minimizes the p-value
  - Advantage: Does not need randomized studies, hence could be based on much larger set of data, even outside molecule program
  - Disadvantage: Why should a value separating best the good versus the bad prognosis sub populations also define best the population who most benefit from a new therapy?

Frequent requirements and solutions
Enrichment definition

- Two main questions:
  - Does BM have any predictive effect?
  - In case of continuous BM: What is the cut off leading to optimal BM effect?
- Predictive effect:
  - ***Qualitative***: Treatment works in BM+ but not in BM-
    (Example: Herceptin and Her2)
  - ***Quantitative***: Treatment works better in BM+ but still works in BM-
    (Example: Iressa, Tarceva and EGFR)

Frequent requirements and solutions
Enrichment definition

- Qualitative and quantitative effects are not the same:
  - For qualitative effects: Defines enrichment population
  - For quantitative effects: Target still all comer population but BM can define population of still unmet medical need
- Important not only to determine predictiveness
  - Regulatory requests to sample large amount of BM- patients data as well likely to differentiate between qualitative and quantitative BM

Frequent requirements and solutions
Enrichment definition

- First question to be answered: is there a (qualitative) predictive effect between treatment and BM?
- Two hypothesis:
  - Test for treatment difference in all comers
  - Test for treatment differences in BM+ patients
  - Both tests correlated usually by fraction of information like in sequential analyses
- How to make this decision process efficient?

Frequent requirements and solutions
Enrichment strategies

Two possibilities:

•Selection strategy:

- Select one hypotheses and/or population before final analysis

- Adjustment strategy:
  - Test both hypotheses at final analysis and adjust for multiplicity

Frequent requirements and solutions
Enrichment strategies

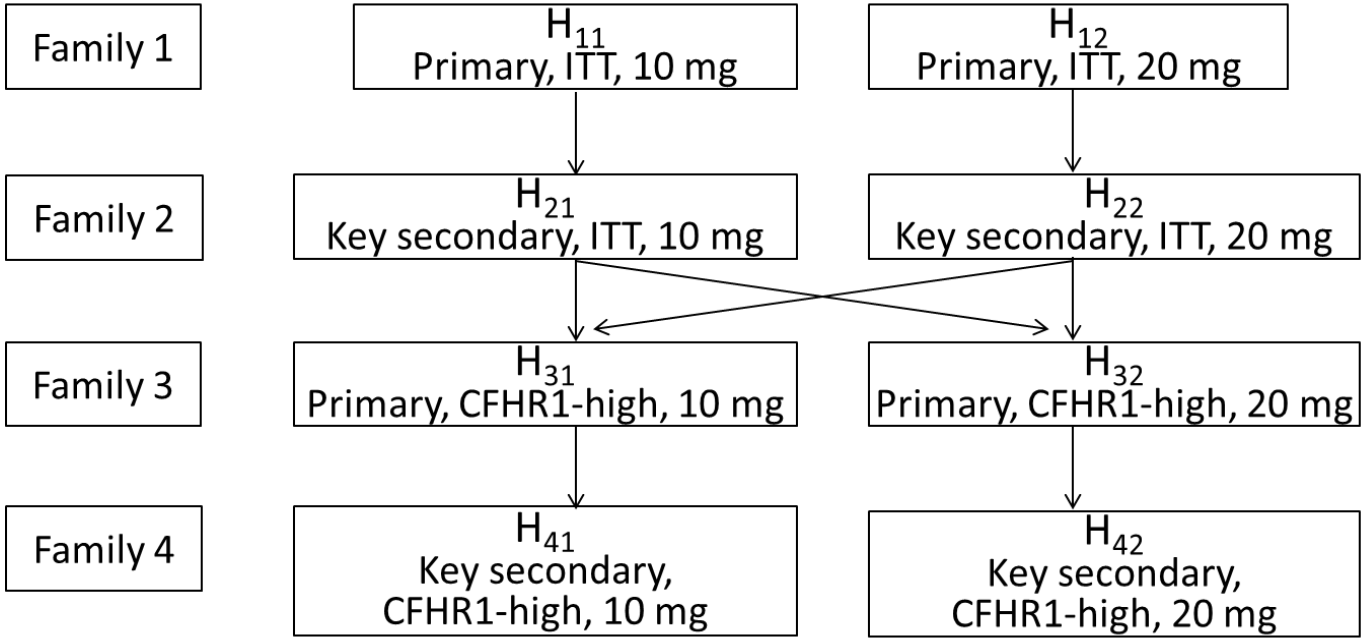
- Selection strategy:
  - A: Make enrichment decision based on phase I data
  - B: Make decision based on phase II data before start of phase III
  - C: Run phase II and III in parallel with the assumption that phase II matures first. Phase II results will then determine primary hypothesis in phase III (protocol amendment before phase 3 code break)
- Both cases:
  - Either inefficient and not using all data or
  - High risk for false decision based on limited data

Frequent requirements and solutions
Enrichment strategies

- Adjustment strategy:
  - Do hierarchical design (test first all comers, then enriched population)
  - Use for example Bonferroni adjustment
  - Use more complex adjustment methods for multiplicity based on correlation between the two tests
  - Use mixture gatekeeping
  - Frequent idea: BM hypothesis as de-risking step
- Adjustment strategy potentially inefficient primarily because of enrolling patients for whom therapy may not work

Frequent requirements and solutions
Example: Mixture gate keeping

- Study with two experimental doses with CHFR1 biomarker enrichment



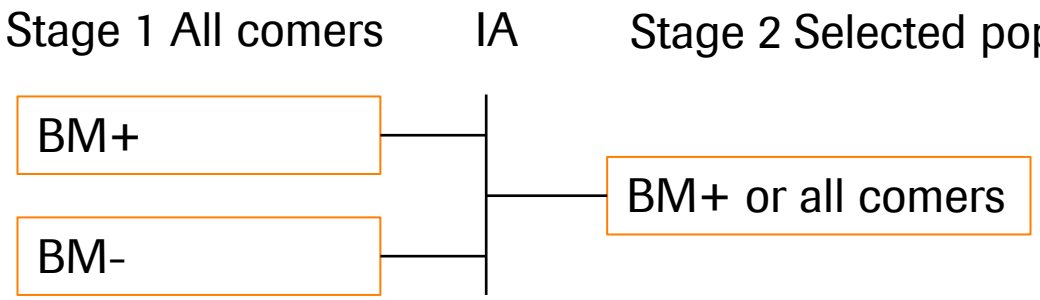
- In this example there is basically not much left for the CHFR1 biomarker

Frequent requirements and solutions
Adjustment strategy

- In summary, strategy should clarify BM question in a way that when overall no effect could be confirmed there is still a chance for the subgroup
  - Bonferroni is conservative but fulfills this
  - Hierarchical designs do not de-risk (when overall is negative then also subgroup is formally negative)
  - Mixture gatekeeping does also not really fulfill this. Usually there is no real power left for BM+ subpopulation

Frequent requirements and solutions
Adaptive designs as an alternative?

- Adaptive designs with a population selection step built in at interim
  - Could use all available information
  - Provide better de-risking step
  - Base enrichment decision on the right size
  - Enroll BM- population only till sufficient information for decision making is available



Frequent requirements and solutions
Adaptive designs as an alternative?

- Adaptive designs with a population selection step built in at interim
  - Design allows to have stage 1 large enough for meaningful decision. Stage 2 may be smaller then
  - Decision criteria and type 1 error control need to be worked out
  - Provides a solution in case of one biomarker hypothesis but likely not when there are more
  - BM- data at interim may be sufficient for overall evaluation
- We need to investigate in which case what design is optimal
  - Not expected that all methods will be equally good or one will be the best in all cases
  - We need criteria to recommend when what should be used

Frequent requirements and solutions
Role of biology

- Biological plausibility for a biomarker effect important
  - Biological plausibility additional information which can help ruling out false positive
    Important: Pre-specification of primary hypothesis or order of plausibility
    (Because we always find some explanation otherwise)
  - Biological plausibility important for exploratory BM analyses but also else. Some biomarker only based on biology (for example in infectious diseases)
  - Generally: The more we understand the biology the less we need to sample data on BM- patients

Frequent requirements and solutions
Role of biology: Prognostic versus predictive effect?

- Basic idea:
  - Patients with worse prognosis could be identified by BM
  - BM is causative for worse prognosis
  - Treatment affecting biomarker (e.g. Blocking BM in patients with high expression) improves prognosis
- Examples:
  - Where it works: Her2 overexpression and Trastuzumab
  - Where it does not work: Low hemoglobin and erythropoietin stimulating therapies
- Be cautious in situations with no overall benefit but a positive effect in the BM+ subpopulation and a negative effect in BM- :
  Unless there is biological plausibility for a negative effect in BM- one should not believe in results

Frequent requirements and solutions
Surrogate endpoints

- Since decision on BM should be done as early as possible availability of a good surrogate endpoint may be an important additional asset
- Decisions on BM development may be done on sparse and early information:
  - Phase I or phase II data
  - Stage 1 data in an adaptive phase 3 study
- Final endpoint not yet observed in sufficient quantities needed for good decision making
- BM decision therefore based on surrogate endpoints

Frequent requirements and solutions
Surrogate endpoints

- Examples
  - Oncology: Final endpoint is overall survival and interim decision endpoints are response rate or progression-free survival
  - BM decision based on response in phase I
  - BM decision based on response or progression free survival in phase II
- Surrogate endpoints correlated with clinical endpoint but predictive property and value of surrogacy often not sufficiently validated, hence using surrogate endpoints comes with additional risks

Summary

- Enrichment designs are a reality already today
- Despite the many challenges we need to focus on personalized medicine for many reasons
- Methods used often simple and appear frequently inefficient but may still turn out to be adequate in certain situations dependent on how much we believe in a BM hypothesis
- Further alternative designs highly needed
- Criteria needed which design is best in which situation

Thank you! &