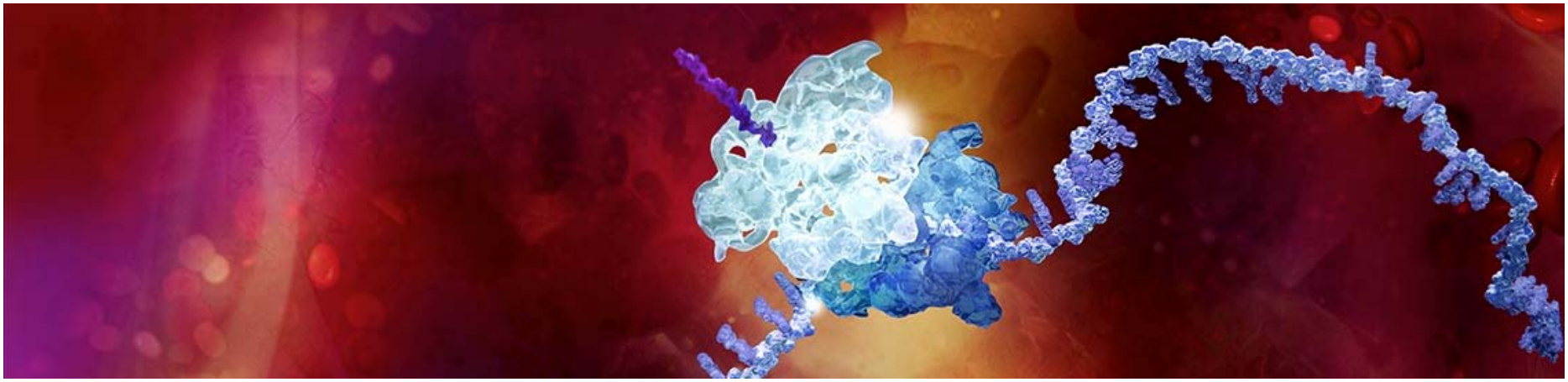
Design of Drug Development Programs with Biomarkers: A stratified medicine approach

Paul Frewer

Translational Statistics: Ideas-Evidence-Innovation-Communication – 30th March 2017



Abstract

In oncology is it becoming increasing common to have targeted treatments based on a patients biomarker status.

The talk will focus on types of designs which could be used in investigating a treatment with this intention.

For example designs allowing assessment of both biomarker positive and negative populations and designs incorporating a number of different investigational medicines targeted at specific patients.

There are advantages and challenges to the designs and we will focus on the statistical considerations to be aware of when developing the clinical plan

Contents

- Need for patient selection
- Types of biomarker
- Enriched/stratified study designs
- Basket/umbrella studies

Need for patient selection

Targeted oncology drugs are designed to interfere with specific molecules that are involved in tumour growth and spread.

There is a biological rationale to restrict development to patients for whom the specific target molecule appears to be playing a key role in their cancer’s growth and survival.

Types of biomarkers

Biomarker

"a biological molecule found in blood, other body fluids, or tissues that is a sign of a normal or abnormal process, or of a condition or disease,"(NCI)

Categorical → Continuous

- Is a specific genetic alteration present: yes or no?*
  - EGFR mutation in NSCLC.
  - BRAF in melanoma.
- A quantitative assay (e.g. IHC) is collapsed into a binary result.
  - Estrogen receptor expression in breast cancer.
  - PD-L1 expression in NSCLC.
- PSA levels in prostate cancer.

*although there will still be sources of variability: heterogeneity within tumour over time, between labs, interpretation, etc.

From continuous to categorical: how do you define a threshold?

PD-L1 expression in NSCLC data from a phase I study of pembrolizumab (Dolled-Filhart et al., 2016)

		RECIST response	
		Yes	No
Percent of cells staining positive	≥50%	19	25
	<50%	8	94

If a ≥50% threshold is used to predict RECIST response then the proportion of "true" responses called correctly is 19/27.

The proportion of "true" non-responses called correctly is 94/119.

50% was chosen because all other thresholds would produce a smaller score when these two proportions are added together.

- No control arm used at this stage: we can't tell from this experiment whether this biomarker is predictive or prognostic.
- Should we be confident moving into phase III with thresholds defined in early stages of development?
- Jiang et al. (2007) suggest defining the threshold adaptively in phase III.

Drug development with a companion diagnostic

RESEARCH PAPER

New Biotechnology • Volume 29, Number 6 • September 2012

ELSEVIER

Research Paper

Practical perspectives of personalized healthcare in oncology

Darren R. Hodgson[1], Robert Wellings[1] and Christopher Harbron[2]

[1] Clinical Biomarker Group, Oncology Therapy Area, Global Medicines Development, Macclesfield, UK SK10 4TG
[2] Discovery Statistics, AstraZeneca, Macclesfield, UK SK10 4TG

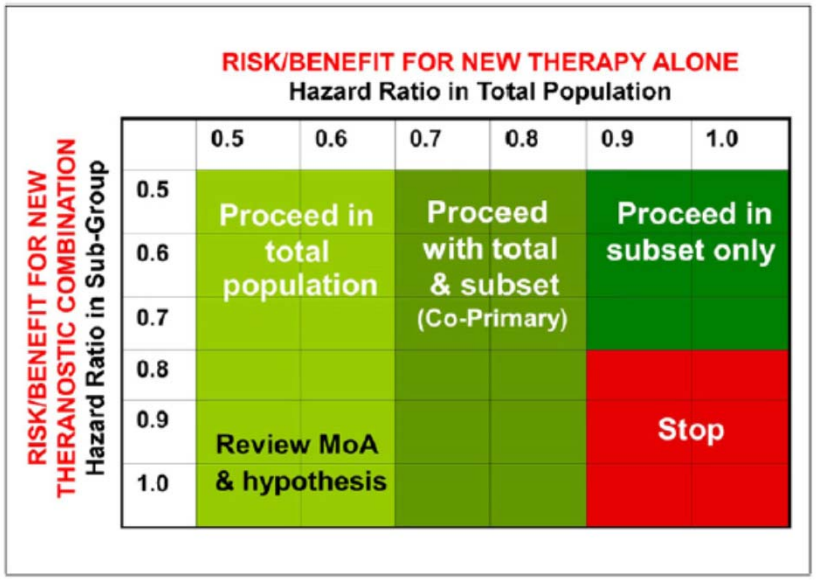


FIGURE 1

Estimating the clinical utility of drug and biomarker. The individual contributions of candidate drug and diagnostic to risk, benefit and value cannot be determined without comparators for standard clinical practice in treatment and diagnosis, that is, without treatment of putative diagnostic 'negative' patients. Numbers for illustration only.

When is it more efficient to proceed in subset only?

Biomarker positive population:

$$\log \hat{\mathrm{HR}}_+ \sim N\left(\ \theta_+\ ,\ \frac{4}{e_+}\ \right)$$

i.e.,

$$\log \hat{\mathrm{HR}}_+ \frac{\sqrt{e_+}}{2} \sim N\left(\ \sqrt{e_+}\frac{\theta_+}{2}\ ,\ 1\ \right)$$

Biomarker negative population:

$$\log \hat{\mathrm{HR}}_- \sim N\left(\ 0\ ,\ \frac{4}{e_-}\ \right)$$

Full population, where p is the proportion biomarker positive and e is the total number of events:

$$p \log \hat{\mathrm{HR}}_+ + (1-p) \log \hat{\mathrm{HR}}_- \sim N\left(\ p\theta_+\ ,\ \frac{4}{e}\ \right)$$

i.e.,

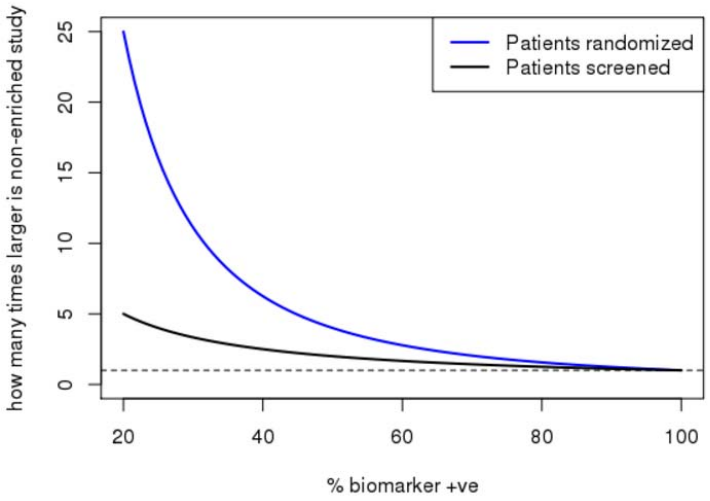
$$\left(p \log \hat{\mathrm{HR}}_+ + (1-p) \log \hat{\mathrm{HR}}_-\right) \frac{\sqrt{e}}{2} \sim N\left(\ p\sqrt{e}\frac{\theta_+}{2}\ ,\ 1\ \right)$$

Therefore the number of events a non-enriched study needs, in order to have the same power as an enriched study with e_+ events, is

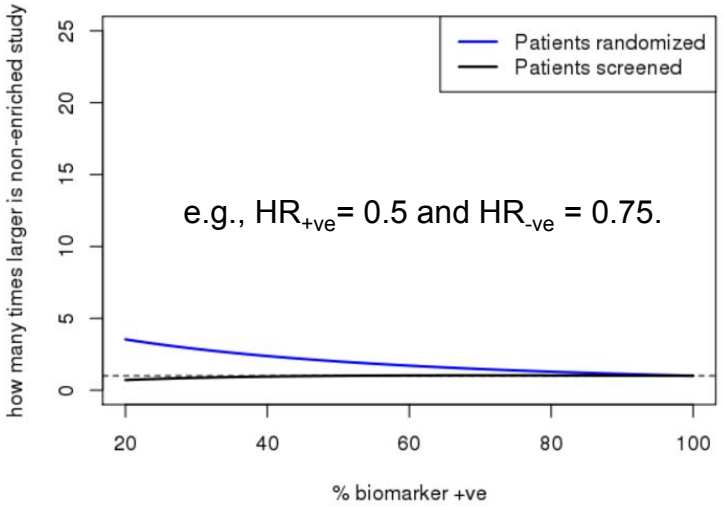
$$e = e_+ / p^2$$

When is it more efficient to proceed in subset only?

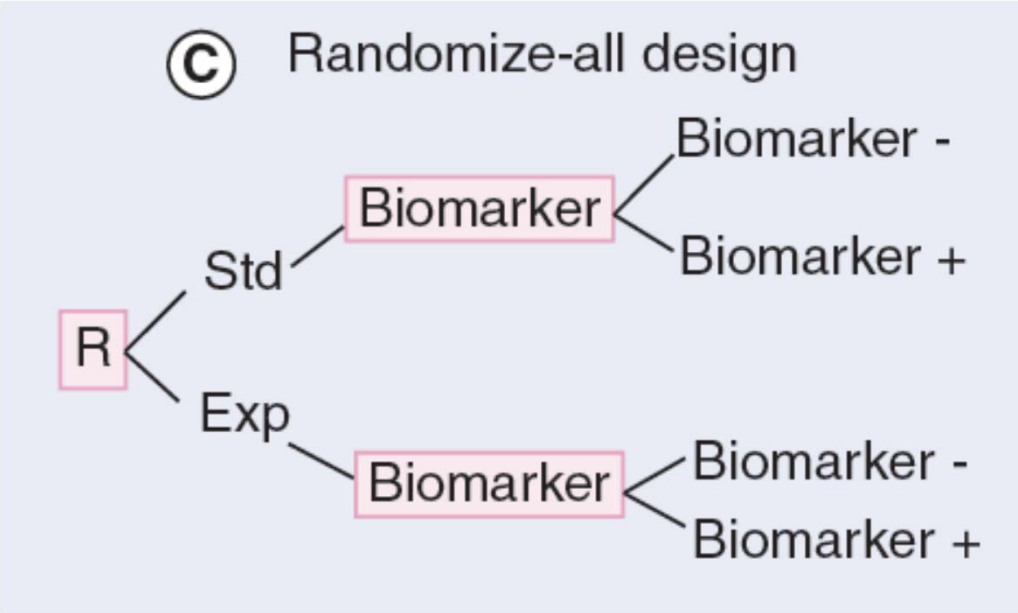
No effect in biomarker negative group



Some effect in biomarker negative group



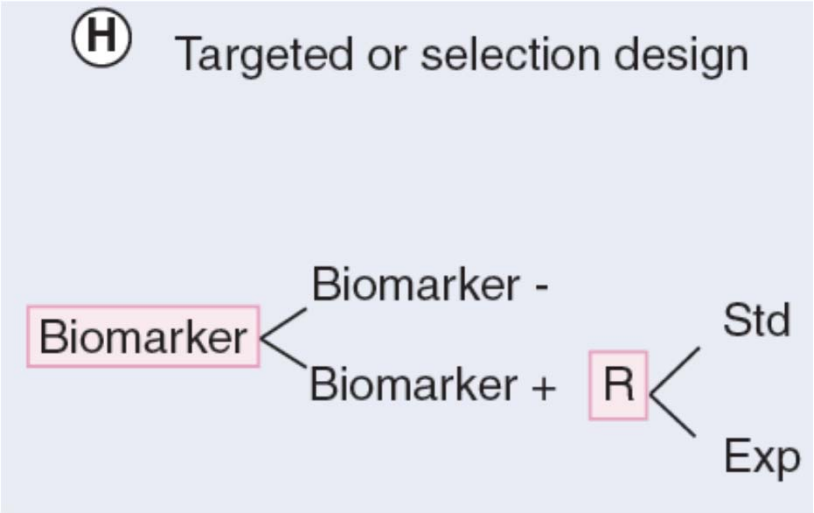
Randomize-all design



Expert Rev. Mol. Diagn. 11(2), 171–182 (2011)

- may have insufficient patients to assess effect in Biomarker +/-
- imbalance in Biomarker +/-

Targeted design

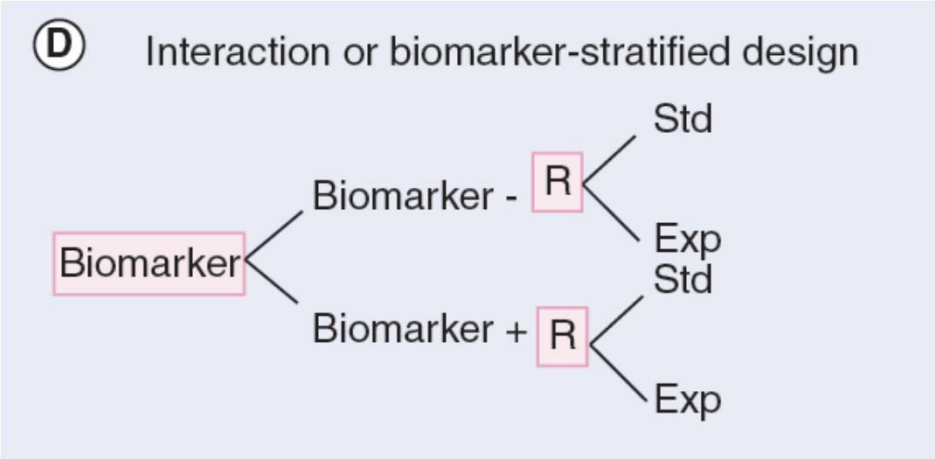


Expert Rev. Mol. Diagn. 11(2), 171–182 (2011)

Vemurafenib and BRAF V600 mutations. Melanoma – the BRIM-3 study (Chapman, 2011).

- 2107 patients screened – 675 patients randomized.
- Vemurafenib vs. dacarbazine.
- OS hazard ratio: 0.23 (0.26,0.55)
- PFS hazard ratio: 0.26 (0.2,0.33)

Stratified design



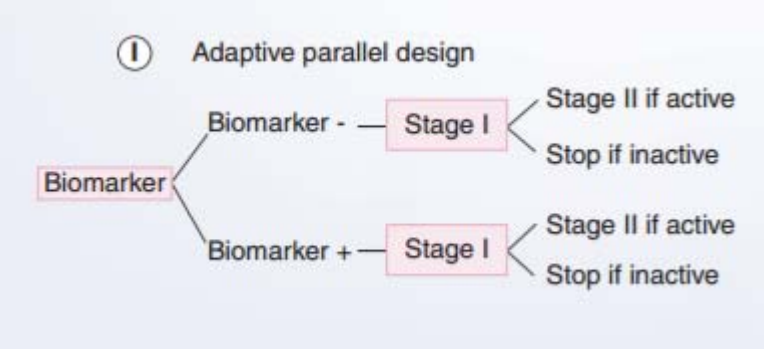
MARVEL: A Phase III Biomarker Validation Study of Second-Line Therapy in Patients With Advanced Non- Small Cell Lung Cancer (NSCLC)

- Used EGFR mutation status as a biomarker.
- Initially planned to enroll 957 patients
- Trial stopped after 23 patients randomized (slow recruitment)
- ClinicalTrials.gov
- NCT00738881

Expert Rev. Mol. Diagn. 11(2), 171–182 (2011)

A strata maybe capped, so the All patient population is enriched for the presence or absence of the biomarker

Adaptive design



Expert Rev. Mol. Diagn. 11(2), 171–182 (2011)

- This strategy has been used in phase II oncology studies at AZ.
- Increased flexibility.
- Analysis more challenging.

MAIN PAPER

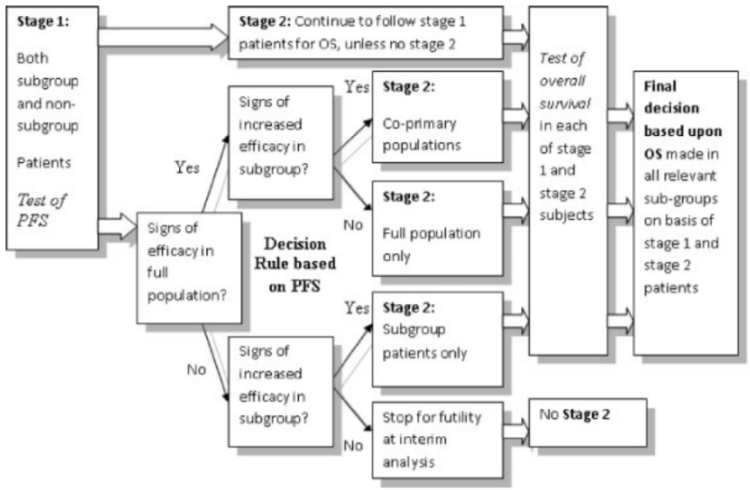
Pharmaceutical Statistics

(wileyonlinelibrary.com) DOI: 10.1002/pst.472

Published online 8 December 2010 in Wiley Online Library

An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints†

Martin Jenkins,[a]* Andrew Stone,[a] and Christopher Jennison[b]



- In an adaptive design you don't have to use the primary endpoint to decide which population to go into at the interim analysis.
- For example, PFS is used to select population, OS is used at final analysis.
- However, not straightforward to control the type I error rate.

Possible structure of Phase 3 trial with BM

Key features of phase III biomarker study designs

Table 2 | Key features of phase III biomarker study designs

Design	Groups evaluated	Advantages	Disadvantages	Recommended use
Enrichment	Biomarker-positive subgroup only	Efficient test of treatment benefit in biomarker-positive population	Provides no evidence of treatment benefit in biomarker-negative patients; Provides no direct evidence (by itself) of clinical utility of biomarker	Biomarkers with very strong credentials; Convincing evidence that the benefit is limited to biomarker-positive patients
Stratified designs				
Subgroup-specific	Biomarker-positive subgroup and biomarker-negative subgroup (sequentially or in parallel)	Provides clear evidence of treatment benefit in the biomarker-positive population and the biomarker-negative population	Has suboptimal power when treatment effect is homogeneous across biomarker subgroups	Biomarkers with strong credentials; Convincing evidence that the treatment is more effective in biomarker-positive than in biomarker-negative patients
Biomarker-positive and overall	Biomarker-positive subgroup and overall population (sequentially or in parallel)	Provides clear evidence of treatment benefit in the biomarker-positive population; Good statistical power when the treatment effect is homogeneous across subgroups	Does not provide a clear assessment of benefit in biomarker-negative patients; Can lead to recommendations for use of an ineffective treatment in biomarker-negative patients	Not recommended
Marker sequential test	Biomarker-positive and biomarker-negative subgroups, and overall population (sequentially)	Provides clear evidence of treatment benefit in the biomarker-positive population and in the biomarker-negative population; Good statistical power when treatment effect is homogeneous across subgroups	Requires a marginal sample size increase compared to subgroup-specific design	Biomarkers with strong credentials; Convincing evidence that the treatment is more effective in biomarker-positive than in biomarker-negative patients
Fallback	Overall population and biomarker-positive subgroup (sequentially)	Enables statistically rigorous evaluation of treatment benefit in the biomarker-positive patients, if no benefit is detected in the overall population	Does not provide clear assessment of benefit in the biomarker-negative patients	Biomarkers with weak credentials; Treatment assumed to probably benefit all patients and/or biomarker-positive group is only assessed if no benefit is shown in overall population

Freidlin, B. & Korn, E. L. (2014) Biomarker enrichment strategies: matching trial design to biomarker credentials, *Nat. Rev. Clin. Oncol.*

Enrichment and stratified designs: discussion

Scientific/Statistical

- Enrichment/targetted design is simple to analyse.
- Stratified design makes it possible to assess the predictive ability of the biomarker.

Practical

- Enrichment studies will have slow enrolment when the prevalence of biomarker +ve population is low.
- Stratified studies can have a similar problem: the biomarker negative population may finish recruitment a lot earlier than the biomarker positive.

These practical issues can be mitigated by testing several targeted agents at the same time in “multi-drug basket” or “umbrella” studies…

Basket studies

Traditional	Basket	Multidrug basket
• 1 site of disease	• Many sites	• Many sites
• 1 { Mutation / Targeted agent	• 1 { Mutation / Targeted agent	• Many { Mutation / Targeted agent
	Vemurafenib and BRAF V600 mutations. Many sites (excluding melanoma).	*NCI Match*

Basket studies: design

Typically single arm studies, with primary endpoint of objective response rate and/or duration of response.

Analysis might (or might not) lump together patients with different sites of disease.

Vemurafenib and BRAF (non - melanoma).

- 6 substudies defined by site (NSCLC, breast, etc.)
- In each substudy:
  - Adaptive Simon's 2-stage design.
  - Between 7 and 19 patients.
  - Aim is to reject a null hypothesis that ORR is 15%.

NCI Match

- At least 22 substudies defined by drug target.
- In each substudy:
  - Single-stage design.
  - 31 evaluable subjects.
  - Aim is to reject null hypothesis that ORR is 5%.

Umbrella studies

Enrichment →

- 1 site
- 1 { Mutation; Targeted agent vs. control

Vemurafenib and BRAF V600 mutations. Melanoma.

Umbrella

- 1 site
- Many { Mutation; Targeted agent vs. control

Lung MAP

Umbrella studies: design

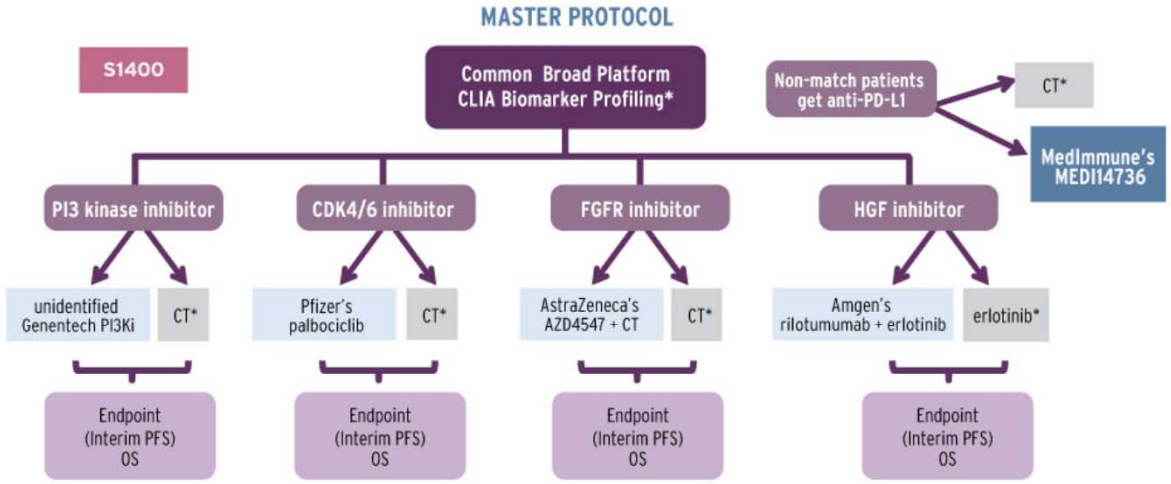
- Typically, a randomized comparison.
- Primary endpoint PFS or OS.

Lung MAP

- 5 substudies based on mutation status.
- In each substudy:
  - Phase II portion has a target PFS hazard ratio of 0.5
  - Go/No-go performed after 55 events.
  - Phase III portion has a target OS hazard ratio of 0.67.
  - Final analysis performed after 256 deaths. (90% power with 2.5% type I error rate).

Mutation	Prevalence
PI3K	6-8%
CDK4/6	12%
FGFR	9%
C-MET	16%
Nonmatch	56%

Lung MAP protocol



CT – chemotherapy (docetaxel or gemcitabine)

* – Archival formaldehyde fixed-paraffin embedded tumor, fresh core needle biopsy if needed

Specific markers for screening

PI3K – PIK3CA mutation
FGFR – FGFR amplication, mutation, fusion
CDK4/6 – CCND1, Cdk6 amplification, CDKN2 deletion and mutation
HGF – c-Met expression

Basket and umbrella studies: discussion

- Just like enrichment designs, they don't directly address the predictive ability of a biomarker.
- Many studies so far have used a default design for each sub-study, with limited statistical justification.
- Should analysis lump together patients who have the same mutation but different sites of disease? Or should separate analyses be performed? Ongoing methodology work here: LeBlanc (2009), Simon (2014).
- Mismatches in recruitment rates between substudies is not eliminated and is still difficult in practice.

References

Dolled-Filhart, Marisa, et al. "Development of a Companion Diagnostic for Pembrolizumab in Non–Small Cell Lung Cancer Using Immunohistochemistry for Programmed Death Ligand-1." *Archives of Pathology & Laboratory Medicine* 140.11 (2016): 1243-1249.

Jiang, Wenyu, Boris Freidlin, and Richard Simon. "Biomarker-adaptive threshold design: a procedure for evaluating treatment with possible biomarker-defined subset effect." *Journal of the National Cancer Institute* 99.13 (2007): 1036-1043.

Jenkins, Martin, Andrew Stone, and Christopher Jennison. "An adaptive seamless phase II/III design for oncology trials with subpopulation selection using correlated survival endpoints." *Pharmaceutical statistics* 10.4 (2011): 347-356.

Hodgson, Darren R., Robert Wellings, and Christopher Harbron. "Practical perspectives of personalized healthcare in oncology." *New biotechnology* 29.6 (2012): 656-664.

Simon, Richard. "Biomarker based clinical trial design." *Chinese clinical oncology* 3.3 (2014).

Buyse, Marc, et al. "Integrating biomarkers in clinical trials." *Expert review of molecular diagnostics* 11.2 (2011): 171-182.

Chapman, Paul B., et al. "Improved survival with vemurafenib in melanoma with BRAF V600E mutation." *n Engl J Med* 2011.364 (2011): 2507-2516.

LeBlanc, Michael, Cathryn Rankin, and John Crowley. "Multiple histology phase II trials." *Clinical Cancer Research* 15.13 (2009): 4256-4262.

Simon, Richard. "Biomarker based clinical trial design." *Chinese clinical oncology* 3.3 (2014).

Freidlin, B. & Korn, E. L. (2014) Biomarker enrichment strategies: matching trial design to biomarker credentials, *Nat. Rev. Clin. Oncol.*

Acknowledgments

Dominic Magirr for his assistance in developing the slides

Amir Handzel

Questions?

Confidentiality Notice