Design Concept for a Confirmatory Basket Trial

Robert A. Beckman, MD[1] and Cong Chen, PhD[2]

[1]Professor of Oncology & of Biostatistics, Bioinformatics, and Biomathematics

Lombardi Comprehensive Cancer Center and Innovation Center for Biomedical Informatics

Georgetown University Medical Center

Founder and Chief Scientific Officer, Oncomind, LLC

[2]Executive Director, Biostatistics and Research Decision Sciences, Merck & Co, Inc.

Deecember 4, 2018

Acknowledgements

- Co-authors on the initial design:
  - **Cong Chen—led group; co-led concept development; led all statistical and simulation work for initial design**
  - Zoran Antonijevic, Amgen
  - Rasika Kalamegham, Genentech
- Pathway design subgroup, additional members:
  - Christine Gausse, Merck
  - Sebastian Jobjorrnsson, Chalmers
  - Lingyun Liu, Cytel
  - Sammy Yuan, Merck
  - Yi (Joey) Zhou, Ultragenyx
  - Advisor: Sue-Jane Wang
- Pathway design subgroup is one of 5 working subgroups of the **DIA Small Populations Workstream**, a group of 50 statisticians and clinicians from industry, academia, and national health authorities (FDA and EMEA)
- Small populations workstream is part of **DIA Adaptive Design Scientific Working Group (ADSWG),** a group of > 200 statisticians and clinicians from industry, academia, and national health authorities (FDA and EMEA)

Joining DIA working groups

- Adaptive design scientific working group: email me: robert.beckman@georgetown.edu or eniac1915@gmail.com
- Bayesian scientific working group: email Fanni Natanegara: natanegara_fanni@lilly.com

Non-Oncology Rare Diseases

- Estimates of rare diseases
  - Up to 7000 different rare diseases
  - Estimated 30 million sufferers in US alone
  - About half are children
  - Many of these diseases are progressive, debilitating, and lethal

Small Populations Within A Common Disease, Cancer

- The increasing discovery of molecular subtypes of cancer leads to small subgroups that actually correspond to orphan or “niche” indications, even within larger tumor types
- Enrolling enough patients for confirmatory trials in these indications may be challenging.
- The shift to a molecular view of cancer requires a corresponding paradigm shift in drug development approaches
- Exclusive use of “one indication at a time” approaches will not be sustainable

Approaches to development based on predictive biomarkers

- Optimized co-development of a single drug and its companion diagnostic
  - Gives a clear hypothesis and answer and still has a role in selected instances
  - Will be challenging to do in niche indications
- "Umbrella" trials
  - One tumor type with multiple drugs and predictive biomarkers
  - Patients are matched to drugs based on predictive biomarkers
  - Cooperation among multiple sponsors
  - Examples: BATTLE, I-SPY, Lung-MAP

Basket Trials

- Multiple tumor types with one drug and predictive biomarker
- Evaluation often based on pooled analysis
  - In some designs, pooling can be partial, based on Bayesian hierarchical model. Degree of pooling can be adjusted based on data
  - In some designs, indications are considered individually. Basket is then more of an operational tactic
- Premise is that molecular subtype is more fundamental than histology
- Can be single sponsor or consortium
- Opportunity for multiple indications for the sample size of one: dramatic potential patient and resource savings

Agenda

- Introduction
- General Design Concept for a Confirmatory Basket Trial
- Challenges and Recommendations for Overcoming Them
- Performance Simulations and Design Considerations
- Conclusions

The Original Basket: Imatinib B2225

186 patients with

40 different malignancies with known genomic mechanisms of activation of imatinib target kinases

KIT, PDGFRA or PDGFRB

Imatinib 400- 800 mg BID primary Endpoint ORR

Synovial Sarcoma	Aggressive Fibromatosis	Dermato-fibrosarcoma Protuberans	Aggressive Systemic mastocytosis	Hyper-eosinophilic syndrome	Myelo-proliferative disorder
1/16 (6%)	2/20 (10%)	10/12 (83%)	1/5 (20%)	6/14 (43%)	4/7 (58%)

13 centers in consortium: North America, Europe, Australia

Lead to supplemental indications for these 4 subsets after pooling with other trials and case reports

Blumenthal, Innovative trial designs to accelerate the availability of highly effective anti-cancer therapies: an FDA perspective, AACR 2014

Pembrolizumab Approval in MSI-High in Solid Tumors

- MSI High is a DNA repair defect that leads to an increased mutational burden within affected tumor cells
- Enhanced mutational burden is believed to create neoantigens, leading to enhanced immunogenicity and greater chance of benefit from a PD-1/PD-L1 antagonist
- 5 single arm studies were pooled, 149 total patients
- Primary endpoint was ORR: 39.6%, 78.8% duration > 6 months
- Most patients had colorectal cancer
- 10 other solid tumor types had between 1 and 5 responses
- Approval was for any solid tumor that was MSI-High

Features of These Designs

- A similar design to Imatinib B2225 was endorsed at a Brookings/Friends Conference in 2011
- Common features:
  - Exploratory and opportunistic in nature
  - Single-arm trials with ORR as primary endpoint
  - Often intend to use pooled population for primary analysis to gain broader indication across tumor types (individual tumor type is not adequately powered)
  - Involve ***possibly*** transformative medicines in patients with great unmet need and ***seemingly*** exceptionally strong scientific rationale

Issues

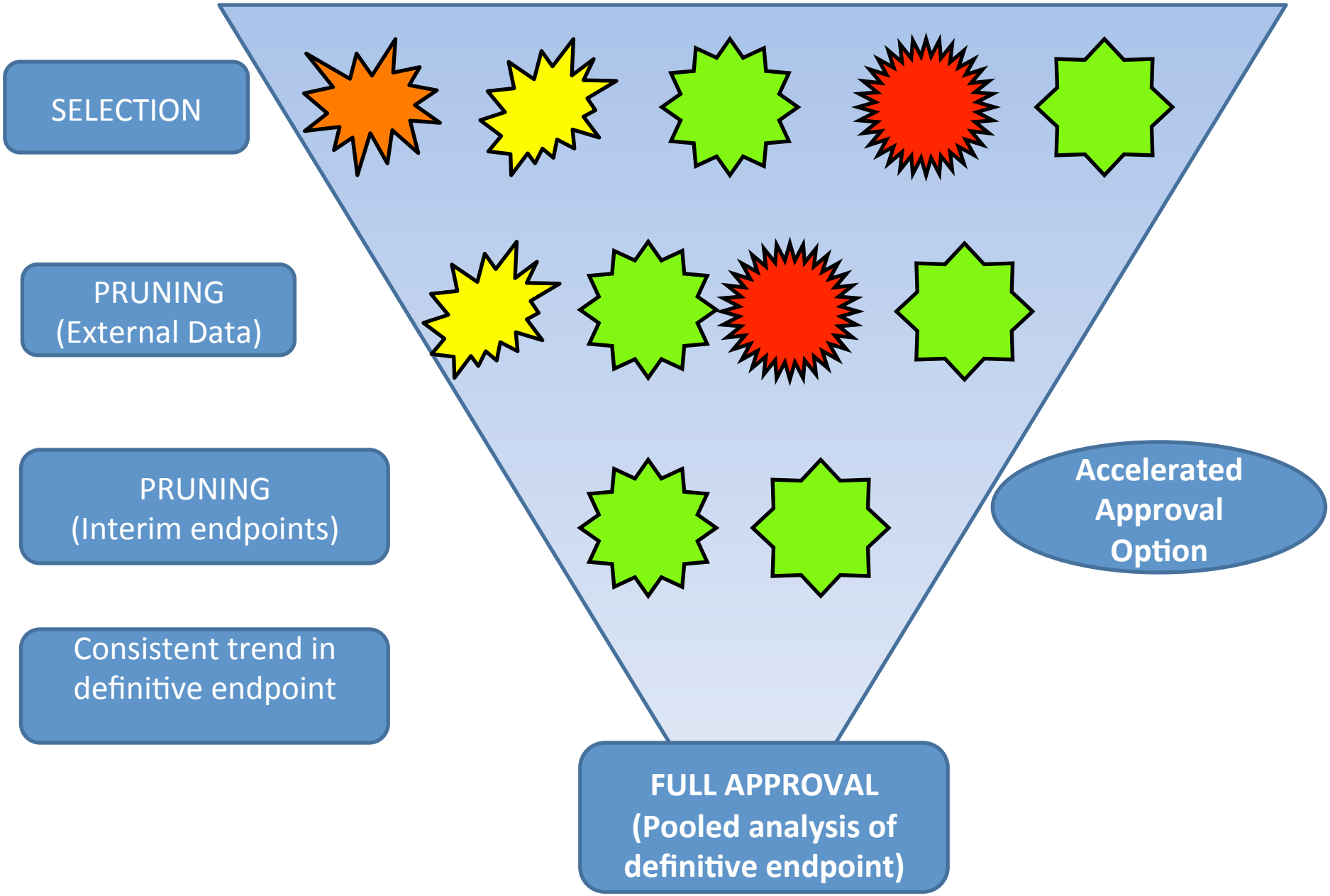
- Clinical data to support pooling may be limited, and treatment effect may differ between tumor types
  - Vemurafenib works in melanoma with BRAF V600E mutation but not colorectal cancer with same mutation
- Not all drugs hoped to be transformational live up to this promise
- Response rate may not predict overall survival
- Single arm trials are subject to patient selection bias
- Predictive effect of a biomarker is confounded with the prognostic value which is often unknown
- Health authorities can be non-committal upfront

DIA Small Population Pathway Subteam

- Can we develop a generalizable confirmatory basket design concept with statistical rigor?
  - Applicable not only to exceptional cases, but to all effective medicines in any line of therapy
  - Follow existing accelerated and standard approval pathways to increase potential approvability
- This would have multiple benefits
  - Increase and accelerate access to effective medicines for patients in niche indications
  - Provide sponsors with cost-effective options for development in niche indications
  - Provide health authorities with more robust packages for evaluation of benefit and risk
  - **MOST OF DRUG DEVELOPMENT RESOURCES ARE SPENT IN THE CONFIRMATORY PHASE**

Most of Drug Development Resources Are Spent in the Confirmatory Phase

GENERAL DESIGN CONCEPT


SELECTION
PRUNING
(External Data)
PRUNING
(Interim endpoints)
Consistent trend in
definitive endpoint
Accelerated
Approval
Option
FULL APPROVAL
(Pooled analysis of
definitive endpoint)

Features of the Design (I)

- Tumor histologies are grouped together, each with their own control group (shared control group if common SOC)
- Randomized control is preferred
  - Single arm cohorts with registry controls may be permitted in exceptional circumstances as illustrated by imatinib B225 and others
- In an example of particular interest, each indication cohort is sized for accelerated approval based on a surrogate endpoint such as progression free survival (PFS)
  - This may typically be 25-30% of the size of a Phase 3 study
- In another approach, an interim evaluation of partial information on the definitive endpoint may be used
- Initial indications are carefully selected as one bad indication can spoil the entire pooled result

Features of the Design (II)

- Indications are further "pruned" if unlikely to succeed, based on:
  - External data (maturing definitive endpoint from Phase 2; other data from class)
  - Internal data on surrogate endpoint OR partial information on definitive endpoint
- Sample size of remaining indications may be adjusted based on pruning
- Type I error threshold will be adjusted to control type I error (false positive rate) in the face of pruning
  - Pruning based on **external** data does not incur a statistical penalty
  - Discussed in more detail later in talk
- Study is positive if pooled analysis of remaining indications is positive for the primary definitive endpoint
  - Remaining indications are eligible for full approval in the event of a positive study
  - Full pooling chosen for simplicity
  - Some of the remaining indications may not be approved if they do not show a trend for positive risk benefit as judged by definitive endpoint

CHALLENGES OF BASKET DESIGNS AND RECOMMENDATIONS FOR OVERCOMING THEM

Challenge #1: Having a Control Group

- In some settings, a control group is not ethical
- Resolution: randomized trial should be applied, if and only if:
  - There is clinical equipoise between the two randomized arms
  - Experimental agent is not expected to be transformational, only beneficial
  - There is an SOC for control
  - Example: steroids +/- rituximab for refractory autoimmune diseases
- Current generation of non-randomized basket studies for transformational agents provides SOC baseline for future randomized trials

Challenge 2: Risks of Pooling

- One of more bad indications can lead to a failed study for all indications in a basket
- Histology can affect the validity of a molecular predictive hypothesis, in ways which cannot always be predicted in advance
  - Vemurafenib is effective for BRAF 600E mutant melanoma, but not for analogous colorectal cancer (CRC) tumors
  - This was not predicted in advance but subsequently feedback loops leading to resistance were characterized

Addressing challenge 2

- Basket trials are recommended primarily after there has been a lead indication approved (by optimized conventional methods) which has validated the drug, the predictive biomarker hypothesis, and the companion diagnostic
  - Example, melanoma was lead indication preceding Brookings trial proposal in V600E mutant tumors
- Indications should be carefully selected
- Indications should be pruned in several steps before pooling

Challenge 3: Different Indications May Have Different Endpoints

- Less of an issue for oncology
- Even in auto-immune diseases, generalized interim endpoints can be created across diverse diseases:
  - Interim: improvement (response)
  - Final: time to worsening

Challenge 4: Timescales of endpoint development may differ

- Resolution:
  - what matters is ***relative*** improvement
  - If necessary, TTE data may be normalized to the medians on control arms of the different indications
  - study completes when data is mature on all arms

Challenge 5: SOC may differ between arms

- Resolution:
  - what matters is ***relative*** improvement ***in a redefined disease entity based on a molecular biomarker***
  - *safety must be assessed both as an individual analysis relative to individual control* ***and*** *as a pooled analysis relative to pooled control*
  - *safety data should be available from reference indications and from phase 2 studies*

Challenge 6: Threshold for Approval May Differ Between Arms

- Resolution: study is judged by pooled result of ***relative*** improvement with statistical and clinical significance
  - Thresholds for such criteria are well known

Challenge 7: Clinical validity of the predictive biomarker hypothesis

- The clinical validity of the predictive biomarker can only be verified by inclusion of “biomarker negative” patients in the confirmatory study
- Addressing the challenge
  - Recommend a smaller pooled, stratified cohort for biomarker negative patients, powered on surrogate endpoint
    - Would need to expand the biomarker negative cohort (to evaluate definitive endpoint) if surrogate endpoint shows possible benefit
  - Prior evidence should permit this if:
    - An approved lead indication has already provided clinical evidence for the predictive biomarker hypothesis
    - Prior phase 2 studies support the predictive biomarker hypothesis in other indications

Challenge 8: Adjusting for Pruning

- Pruning indications that are doing poorly on surrogate endpoints may be seen as cherry picking
  - This can inflate the false positive rate, an effect termed "random high bias"
- Addressing the challenge:
  - Emphasize use of **external data**, especially maturing Phase 2 studies, for pruning
    - Pruning with external data does not incur a penalty for random high bias
  - Apply statistical penalty for control of type I error when applying pruning using **internal data**
    - Methods for calculating the penalty are described in stat methods papers (see key references)
    - Rules for applying penalty must be prospective
    - Penalty is not large enough to offset advantages of design

Type I error control under global null hypothesis

- k tumor indications each with sample size of *N* and all with 1:1 randomization
- An interim analysis is conducted at information fraction *t* for each tumor indication and a tumor will not be included in the pooled analysis if p-value>α_t
- The pooled analysis will be conducted at α^* so that the overall Type I error is controlled at α when there is no treatment effect for any tumor (H0)
- What is α^*?

Solving for adjusted alpha (α^*)

- Let Y_{i1} be the test statistics based on information fraction t, and Y_{i2} be the test statistics based on the final analysis of data in the i-th cohort (i=1, 2,…,k)
- Suppose that m cohorts are included in the final analysis ($m \geq 1$), and let V_m be the corresponding test statistics. The probability of a positive outcome in pooled analysis is

$$Q_0(\alpha^*|\alpha_t, m) = \Pr_{H_0}(\cap\{Y_{i1} > Z_{1-\alpha_t} \text{ for } i=1,\ldots,m\}, \cap\{Y_{j1} < Z_{1-\alpha_t} \text{ for } j=m+1,\ldots k\}, V_m > Z_{1-\alpha^*})$$

or $$Q_0(\alpha^*|\alpha_t, m) = \Pr_{H_0}(\cap\{Y_{i1} > Z_{1-\alpha_t} \text{ for } i=1,\ldots,m\}, V_m > Z_{1-\alpha^*})(1-\alpha_t)^{(k-m)}$$

- α^* is solved from below where c(k, m) = k!/((k-m)!m!)

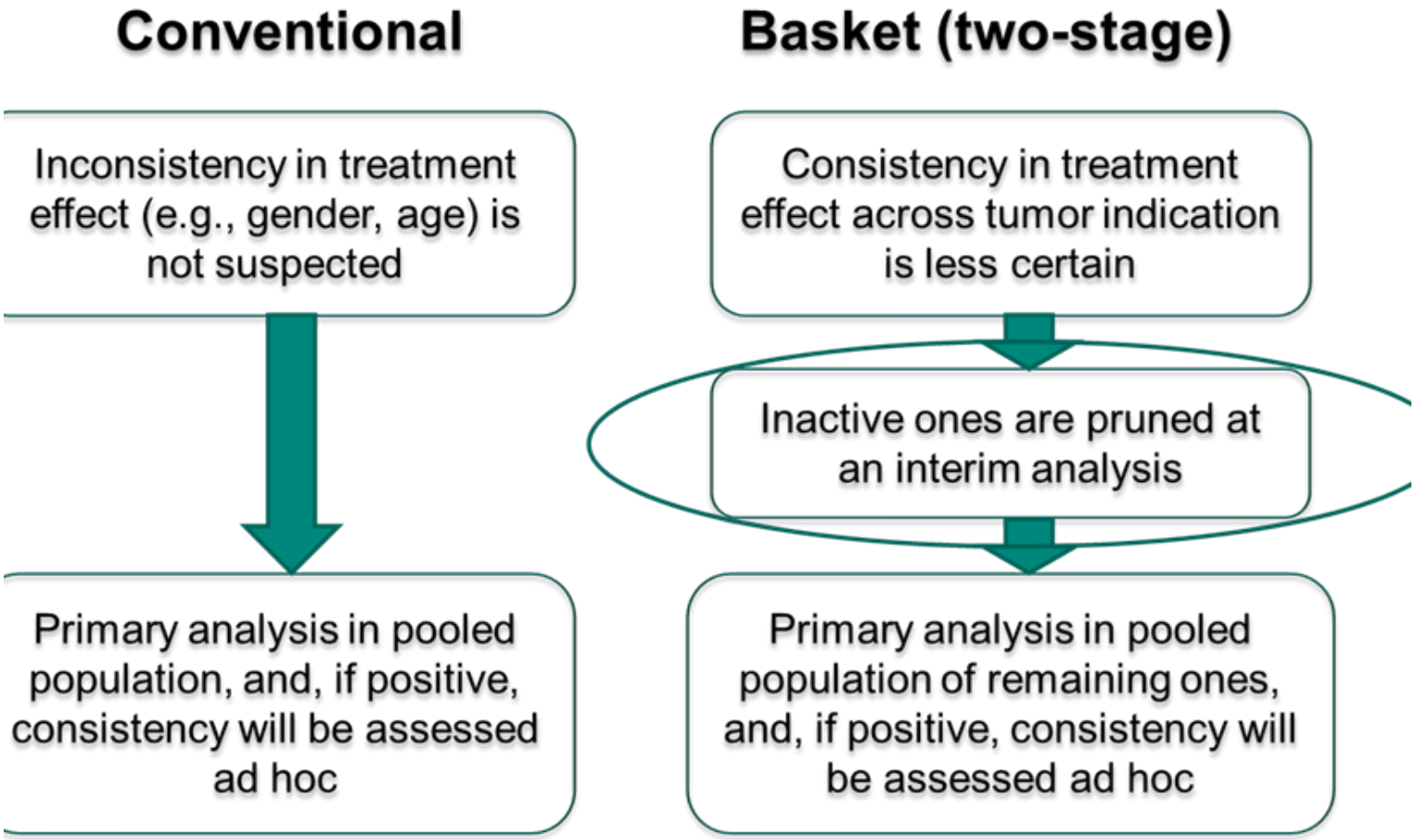
$$\sum_{m=1}^{k} c(k,m) Q_0(\alpha^*|\, \alpha_t, m) = \alpha$$

Challenge #9: Strong Control of FWER

- This problem is still open
- Challenge:
  - One or more strongly positive indications can drive an overall pooled positive result and negative indications are carried along
  - Simulation involves a large number of cases and the degree to which active indications are active affects the results
- A recent MSKCC study* simulated a popular Bayesian basket trial design (using a Bayesian hierarchical model) and found FWER of up to 57%.
  - Authors advocate characterization of FWER by simulation

*Cunanan K et al. Specifying the True- and False-Positive Rates in Basket Trials, *J Clin. Onc. Precision Onc. ,* published online November 3, 2017

Should Basket Trials Control FWER by Indication?



Other FWER Considerations

- A basket trial with k indications replaces k independent trials that collectively would have a family-wise error rate of approximately k * 0.025
- Should we therefore allow approximately k*0.025 for FWER of a basket trial?
- Under would conditions would FDR be a better measure than FWER?

Challenge #10: Availability of tissue

- Tissue sampling and processing are variables that can greatly affect the outcome of a study based on a predictive biomarker
- Basket studies will require cooperation and uniformity across departments organized by histology
- Addressing the challenge:
  - The sponsor must have extensive contact with the pathology department and relevant clinical departments at all investigative sites and provide standard methods for tissue sampling, handling, and processing
  - The sponsor should engage an expert pathologist who is dedicated to training prior to trial start, and troubleshooting during the trial

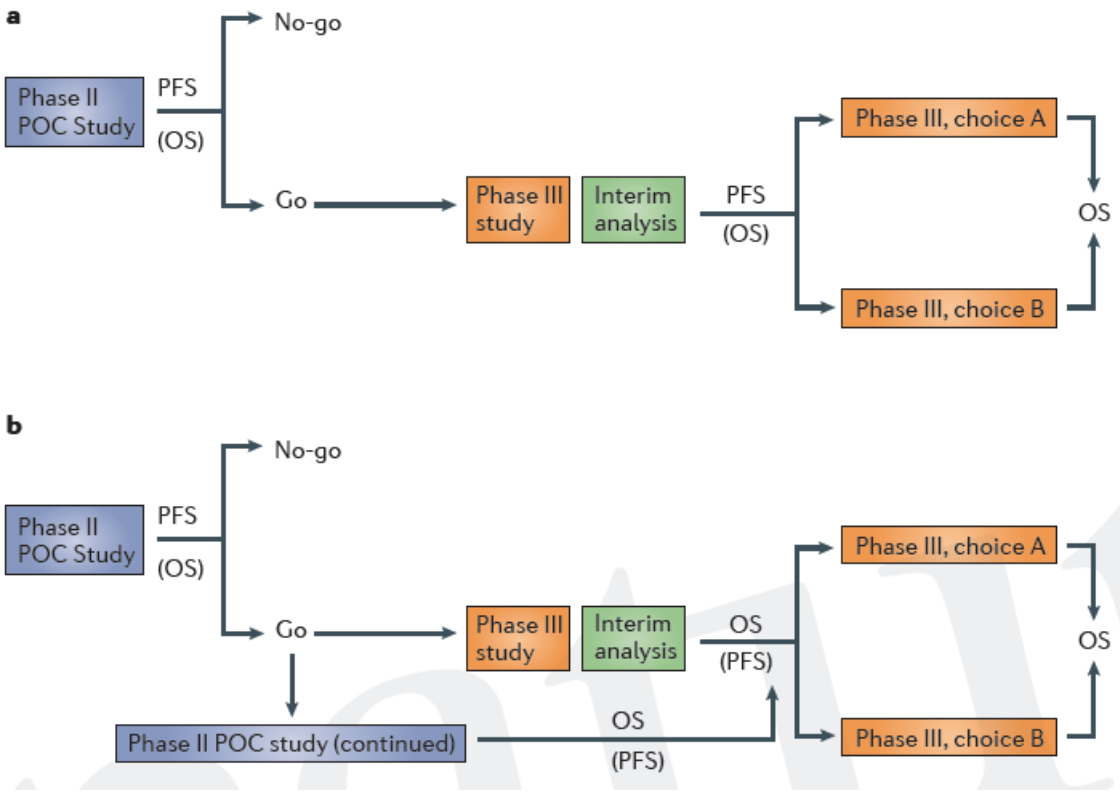
Challenge #11: High Screen Failure Rate

- Pro: patients will have access to tailored therapy
- Con: patient has a high risk of being a screen failure if biomarker positive subgroup is low prevalence
- Addressing the challenge:
  - Study should provide a broad-based test like NGS which will give the patient some guidance on alternative therapies if they are screen failures for basket study

Challenge #12: Interim endpoints may not predict definitive endpoints

- Addressing the challenge:
  - Prefilter indications based on maturing definitve endpoint data from phase 2 or other external data
    - See Figure 2
  - Require consistent trend in definitive endpoint for final full approval

Phase 2 Influencing Phase 3 Adaptation: The Phase 2+ Method



Beckman, R.A., Clark, J. & Chen, C. Integrating predictive biomarkers and classifiers into oncology clinical development programmes. *Nature Reviews Drug Discovery* **10,** 735-748 (2011)

Another Possible Source of External Data

- Real World Data (RWD) from Off-Label Use
- Impact of RWD on basket trial performance is currently under study in a project led by postdoctoral fellow Daphne Guinn

PERFORMANCE SIMULATIONS AND DESIGN CONSIDERATIONS

Comparison of operating characteristics

- k=6 tumor indications with total planned event size (k*N*) ranging from 150-350
  - The true treatment effect is –log(0.6), or hazard ratio of 0.6 in a time-to-event trial
- Pruning occurs at when half of the events have occurred
- Number of active indications (g) with target effect size ranges from 3 to 6, with remaining ones inactive

Study power and sample sizes under different pruning and pooling strategies

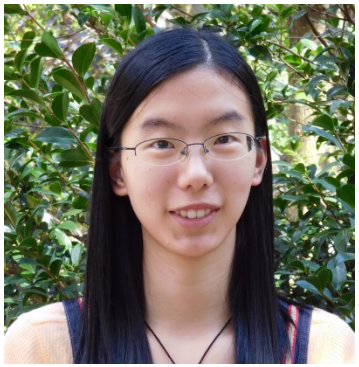
Planned events	Number of active tumors	Power (%) for a positive study				Exp. number of events for pooled population			Exp. number of events for overall study		
		D0	D1	**D2**	D3	D0/D2	D1	D3	D0/D3	D1	D2
200	6	95	85	**95**	93	200	157	179	200	179	221
200	5	85	75	**91**	86	200	144	172	200	172	228
200	4	67	62	**82**	76	200	131	166	200	166	234
200	3	44	45	**68**	61	200	119	159	200	159	240
300	6	99	96	**99**	99	300	254	277	300	277	323
300	5	96	81	**98**	96	300	232	266	300	266	334
300	4	84	81	**94**	91	300	209	255	300	255	345
300	3	60	64	**84**	79	300	187	244	300	244	356

An Application of Special Interest

- A randomized controlled basket trial with 1:1 randomization in 6 tumor indications, each targeting a hazard ratio of 0.5 in PFS with 90% power at 2.5% alpha for global null hypothesis
  - 88 PFS events and 110 patients planned for each indication
  - PFS analysis is conducted when all are enrolled
- D2 is applied to keep total sample size at 660 in pooled population targeting 430 death events
  - The study has ~90% power to detect a hazard ratio of 0.7 in OS at 0.8% alpha (after taking the penalty) assuming $\rho=0.5$
  - Observed hazard ratio ~0.79 or lower for a positive trial in pooled population (vs ~0.84 under D0) for alpha control **under global null**
- Potential to gain approvals in 6 indications based on comparable sample size to a conventional Phase 3 trial

Characterization of Performance Constrained by FWER (ongoing)

- Team includes Yuru Ren, Valeriy Korostyshevskiy, and Sammy Yuan
- Currently studying single TTE endpoint with normally distributed hazard ratios, mean of 1.0 for inactive, 0.7 for active
- Simulate different scenarios of how many indications in basket are inactive. Maximum Type I error (worst case scenario) is FWER
- What power is achievable when FWER must be ≤ k * 0.025?



Current Approaches

- In order to control FWER, we must add an additional post-correction step
- Each indication is tested up to twice individually*
  - at interim information time t [0,1] at significance level alpha-t to govern pruning, AND
  - if part of a successful pool, in a post check at significance level alpha-post

*Beckman R and Loeb LA. Multistage Proofreading in DNA Replication. Quarterly Reviews of Biophysics 26: 225-331 (1993)

Preliminary Results

- k = 6; HR = 0.7, nominal power of pool =95%; t = 0.5, **alpha t = 0.4, alpha post = 0.1:**

6	**0.5276**	**0.9467**	**0.0000**
5	0.5262	0.8954	0.0631
4	0.5079	0.8032	0.1111
3	0.4654	0.6675	0.1432
2	0.3780	0.4739	0.1451
1	0.2316	0.2402	0.0986

- k = 3; HR = 0.7, nominal power of pool =95%; t = 0.5, **alpha t = 0.3, alpha post = 0.1:**

3	0.6940	**0.9534**	**0**
2	0.6815	0.8435	0.0493
1	0.5813	0.5871	0.0731

Future Plans

- Further parameter optimization
- Application of heterogeneity detection methods (Simon)
- Study of application with using surrogate interim endpoint
- Application of RWD to study design

Challenge #14: The Standoff

- Health authorities "understandably" won't commit until given a real example to consider
- Sponsors "understandably" cautious about being first to innovate in confirmatory space
- Resolution:
  - FDA, under **PDUFA VI pilot program**, will be engaging with selected sponsors to bring forward complex innovative designs
  - We must take this risk for our patients

Conclusions

- It is feasible to create a general design concept for a basket study that is suitable for many agents
- Multiple challenges can be addressed with careful planning
- Benefits include:
  - Increased and earlier patient access to targeted therapies for small subgroups
  - Cost-effective methods for sponsors to develop targeted agents in small subgroups
  - More robust datasets for health authorities to assess benefit-risk in these small patient groups

Key References

- Li, Wen, Chen, Cong, Li, Xiaojun, and Beckman, Robert A. Estimation of treatment effect in two-stage confirmatory oncology trials of personalized medicines. Statistics in Medicine, in press (2017).
- Beckman, Robert A., Antonijevic, Zoran, Kalamegham, Rasika, and Cong Chen. Adaptive Design for a Confirmatory Basket Trial in Multiple Tumor Types Based on a Putative Predictive Biomarker. Clinical Pharmacology and Therapeutics, 100: 617-625 (2016).
- Yuan, Shuai S, Chen, Aiying, He, Li, Chen, Cong, Gause, Christine K, and Robert A. Beckman. On Group Sequential Enrichment Design for Basket Trial. Statistics in Biopharmaceutical Research, 8: 293-306 (2016).
- Chen, Cong, Li, Nicole, Yuan, Shuai, Antonijevic, Zoran, Kalamegham, Rasika, and Robert A. Beckman. Statistical Design and Considerations of a Phase 3 Basket Trial for Simultaneous Investigation of Multiple Tumor Types in One Study. Statistics in Biopharmaceutical Research, 8: 248-257 (2016).
- Magnusson BP, Turnbull BW. Group sequential enrichment design incorporating subgroup selection. *Stat Med*. 2013;32(16):2695-2714.
- Heinrich MC, Joensuu H, Demetri GD, Corless CL, Apperley J, Fletcher JA, et al. Phase II, open-label study evaluating the activity of imatinib in treating life-threatening malignancies known to be associated with imatinib-sensitive tyrosine kinases. *Clin Cancer Res*. 2008;14(9):2717-2725.
- Demetri G, Becker R, Woodcock J, Doroshow J, Nisen P, Sommer J. Alternative trial designs based on tumor genetics/pathway characteristics instead of histology. Issue Brief: Conference on Clinical Cancer Research 2011; http://www.focr.org/conference-clinical-cancer-research-2011.
- Antonijevic Z and Beckman RA,, eds. Platform Trials in Drug Development: Umbrella Trials and Basket Trials, CRC Press (Chapman and Hall), 2018,