Statistical Considerations for trials that study multiple indications

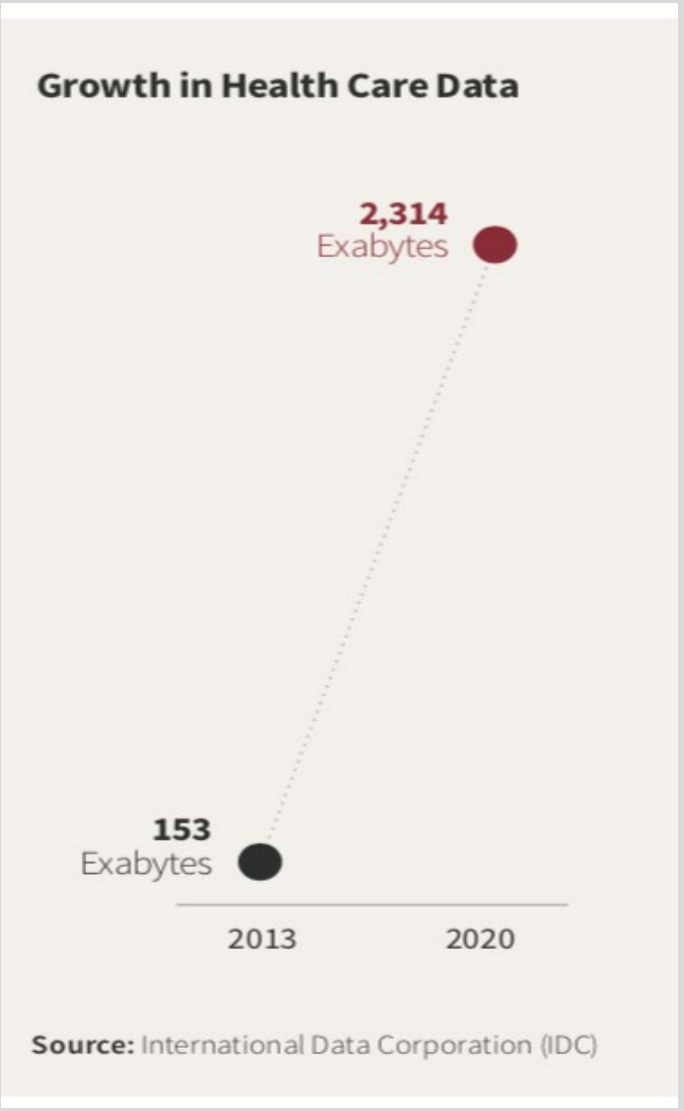
Brian P. Hobbs, PhD
Associate Professor, Dell Medical School
The University of Texas at Austin
brian.hobbs@austin.utexas.edu

DIA BSWG

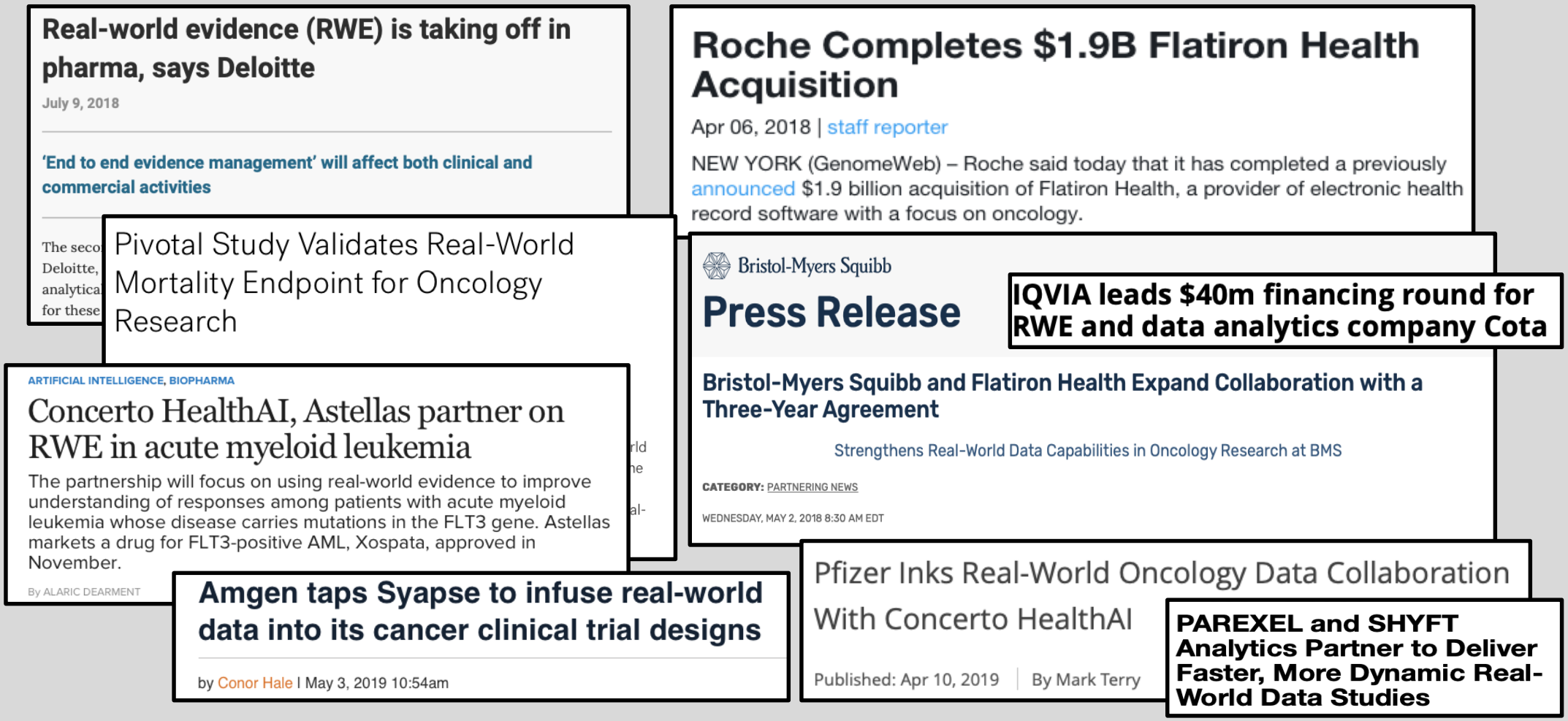
December 18, 2020

The University of Texas at Austin
Dell Medical School

Emerging Big Data Biosphere



Transforming Industry



Regulators – Advocacy for Real-World Evidence

Submitting Documents Using Real-World Data and Real-World Evidence to FDA for Drugs and Biologics

Guidance for Industry

DRAFT GUIDANCE

Accelerating development of scientific evidence for medical products within the existing US regulatory framework

Rachel E. Sherman[1], Kathleen M. Davies[1], Melissa A. Robb[1], Robert M. Califf[1,2]

Growing access to diverse 'real-world' data sources is enabling persistent evidence gaps about the optimal use of medical proc Here, we argue that contrary to widespread impressions, existin sufficient flexibility to accommodate the emerging tools and met

Framework for FDA's Real-World Evidence Program

Real World Evidence

How FDA, Pfizer, and Flatiron Health did it

Approval of Ibrance for men affords a glance at use of real world data

By Paul Goldberg

d a role in FDA's recent decision to expand the indications for Pfizer's drug to include men.

Pivotal Study Validates Real-World Mortality Endpoint for Oncology Research

EW YORK, NY, May 14, 2018

Today, Flatiron Health, a market leader in the curation of regulatory-grade real-world data for cancer research and real-world evidence (RWE) generation, announced the publication of a pivotal study validating the quality of Flatiron's mortality endpoint and illustrating the potential impact for research conducted with contemporary real-world datasets.

FDA cancer office taps Syapse for real-world evidence development

"FDA will work with its stakeholders to understand how RWE can best be used to increase the efficiency of clinical research and answer questions that may not have been answered in the trials that led to the drug approval, for example how a drug works in populations that weren't studied prior to approval."

Janet Woodcock, M.D., Director, CDER

COTA and FDA Partner on Real-World Evidence Program in Breast Cancer

In 2016, Congress passed the 21st Century Cures Act, which, among many things, requires that the U.S. Food and Drug Administration (FDA) take into consideration types of "real-world evidence" when evaluating safety and additional drug indications. It's not, however, absolutely clear what Congress meant by "real-world evidence."

Advocacy for Design Innovation

Regulatory landscape has evolved

- ✓ Approvals small trials w/ refractory subpopulations
- ✓ Advocacy for innovations in design related to precision med
- ✓ Pathway for rapid expansion of phase 1 with seamless design
- ✓ Advocacy for data sharing and leveraging external evidence

The AI hype bubble

AI in Biopharma Slowed by Challenges Involving Data, Corporate Culture

By **Alex Philippidis** - May 15, 2019

As companies address AI bottlenecks, Durvasula of Eli Lilly said, they will be best able to integrate the technologies into their drug discovery and development efforts.

"My hope is that in the next decade, we're going to shift to a compute-first research environment, a model-first research environment, rather than run as many of these experiments as you can, and then do the modeling and figure out what the heck just happened," Durvasula said. "It's got to be a compute-first or a model-first research environment. That requires focusing everybody in the room with all their skills, and with all the multi-domain skills even, on the common purpose, the common scientific question."

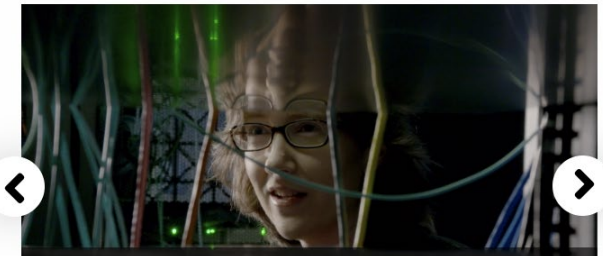
AI in drug discovery is overhyped: examples from AstraZeneca, Harvard, Stanford and Insilico Medicine

In this craze, lots of pharma/biotech companies and investors wonder whether they should jump on the bandwagon in 2018, or wait and see.

Not so elementary, Watson: the roadblocks for AI in pharma

By **Chris Lo**

SHARE



Machine learning systems such as IBM's Watson can fail if the data that 'teaches' them isn't carefully curated. Credit: IBM

RECOMMENDED COMPAN

KENX **Validation Univers**

Knowledge Exchange (KENX) has announc their Validation Univ

ABINGDON **Abingdon Health**

Abingdon Health is a innovative, high-volu mHealth and point-o (POC)...

The AI hype bubble

The attractiveness of the proposition has been borne out in the stacks of pharma and biotech investment that has been flowing towards AI drug discovery tech and machine learning-focused start-ups in the last few years. From Merck's AI partnerships with Numerate and Atomwise to GlaxoSmithKline's $43m collaboration with Exscientia and the rise of AI-centric scientific innovators such as BenevolentAI, pharma AI has become a lucrative business, even before substantial evidence of its impact on drug discovery has been fully explored.

Beyond the “Hype”

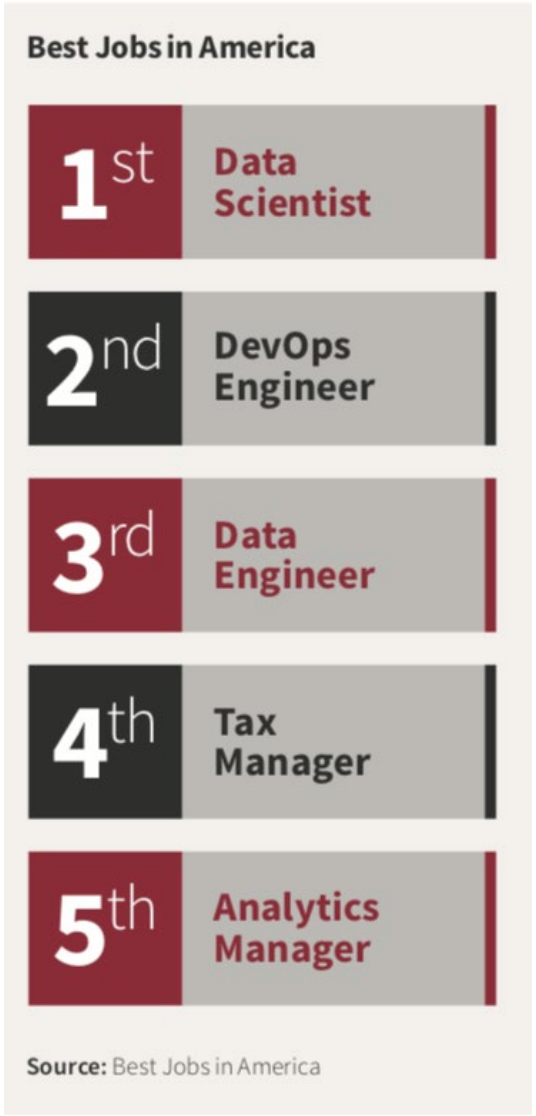
- Data doesn’t do you any good until you can turn it into information, and that is really our challenge.
- … there’s a whole new set of jobs emerging around a health care tech skillset that is very different than it was even just 5 years ago.

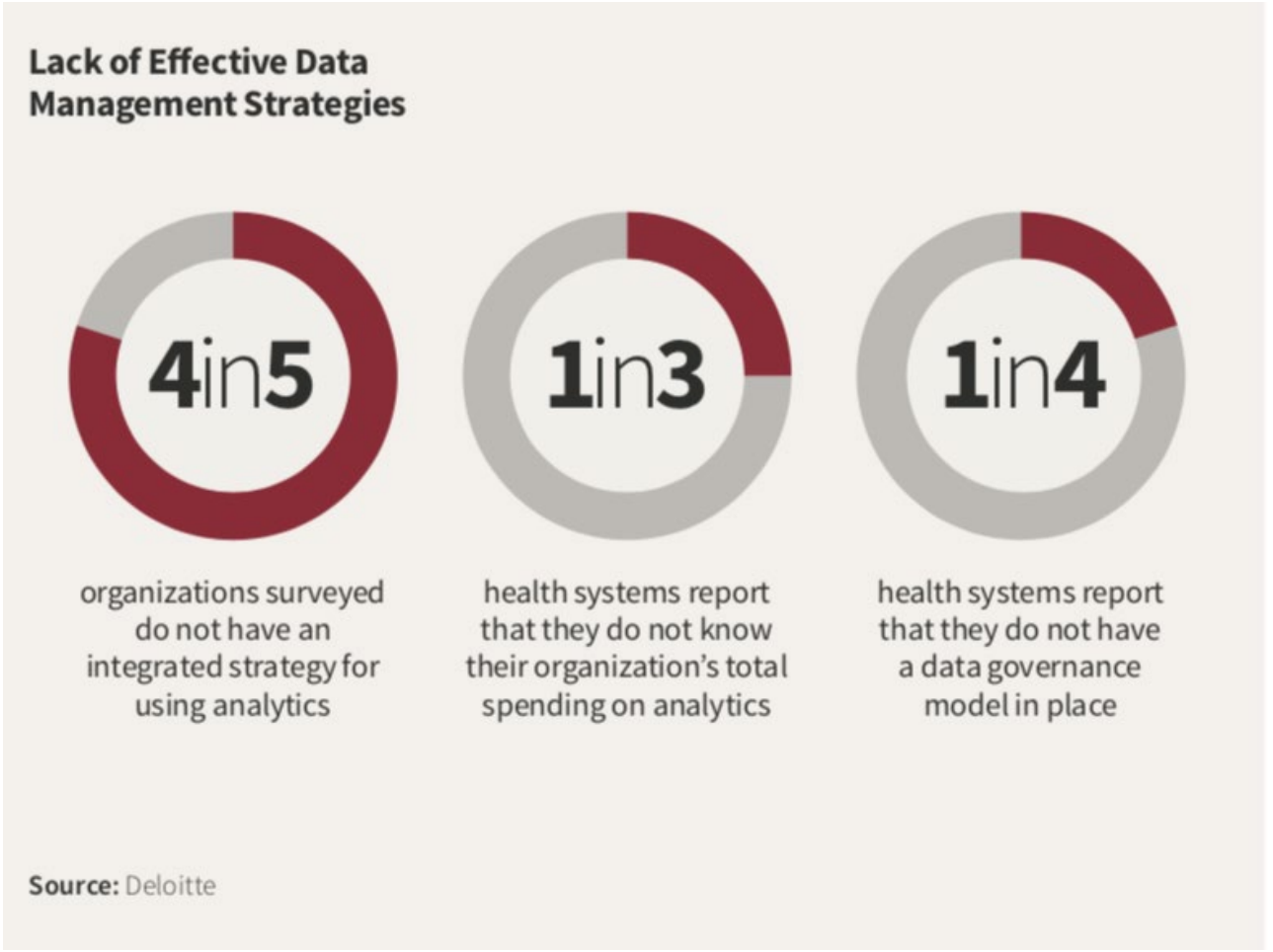
- This health care democratization is characterized by two major factors: the distribution of data and the ability to generate and apply insights at scale.

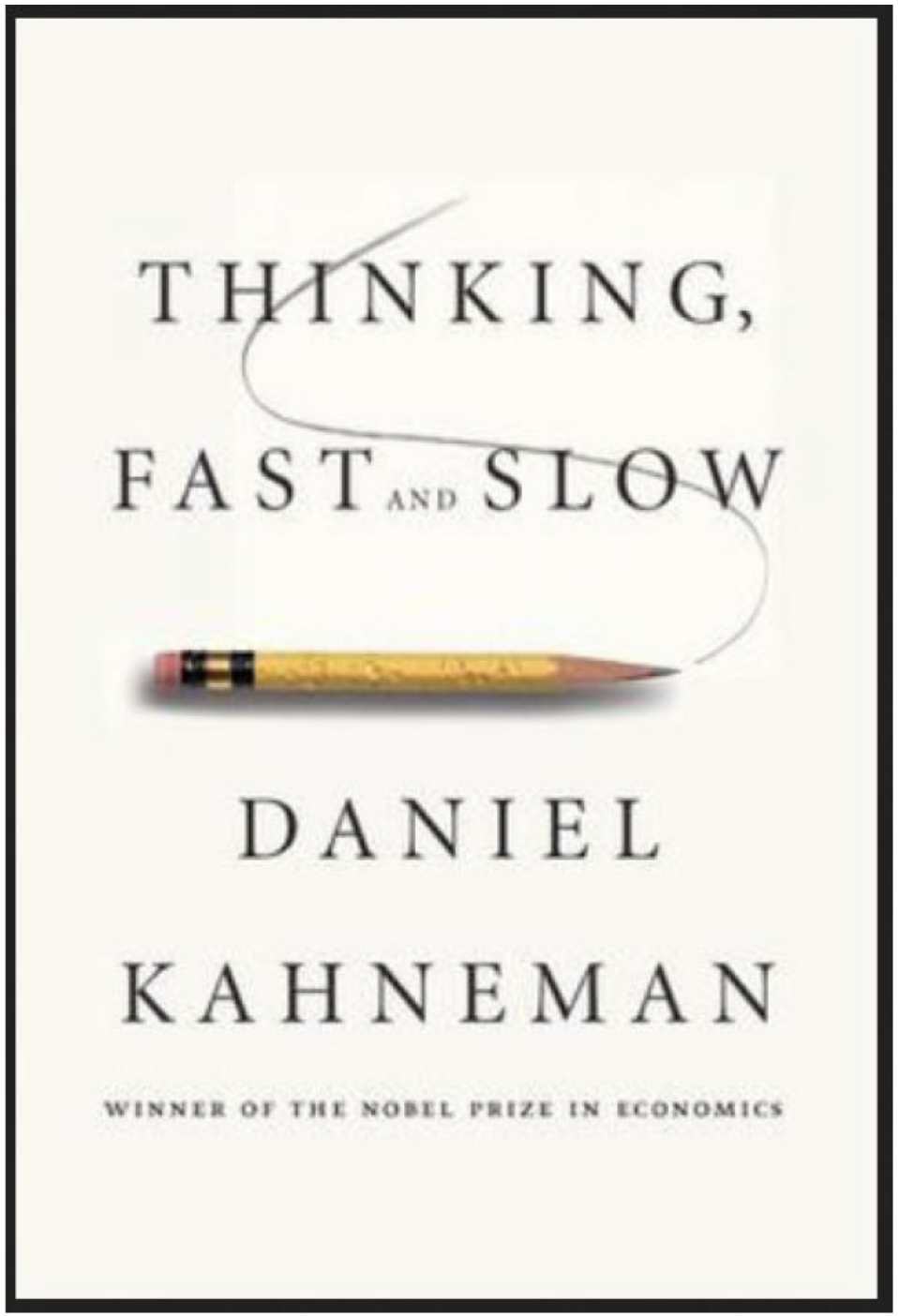
- The biggest problem is that our data are not prepared in a way that allows us to even make sense of it. Once the data are readily analyzable, frankly, the majority of the critical clinical questions can be addressed.

- Amy Abernethy, former Chief Medical Officer/Chief Scientific Officer & SVP Oncology, FlatironHealth; current Principal Deputy Commissioner for Food and Drugs

Expertise Is Key To Avoid Overhyped Claims







Two Modes of Human Cognition

System 1: Fast, automatic, frequent, emotional, unconscious, guided by heuristics

System 2: Slow, effortful, infrequent, logical, conscious

Outline

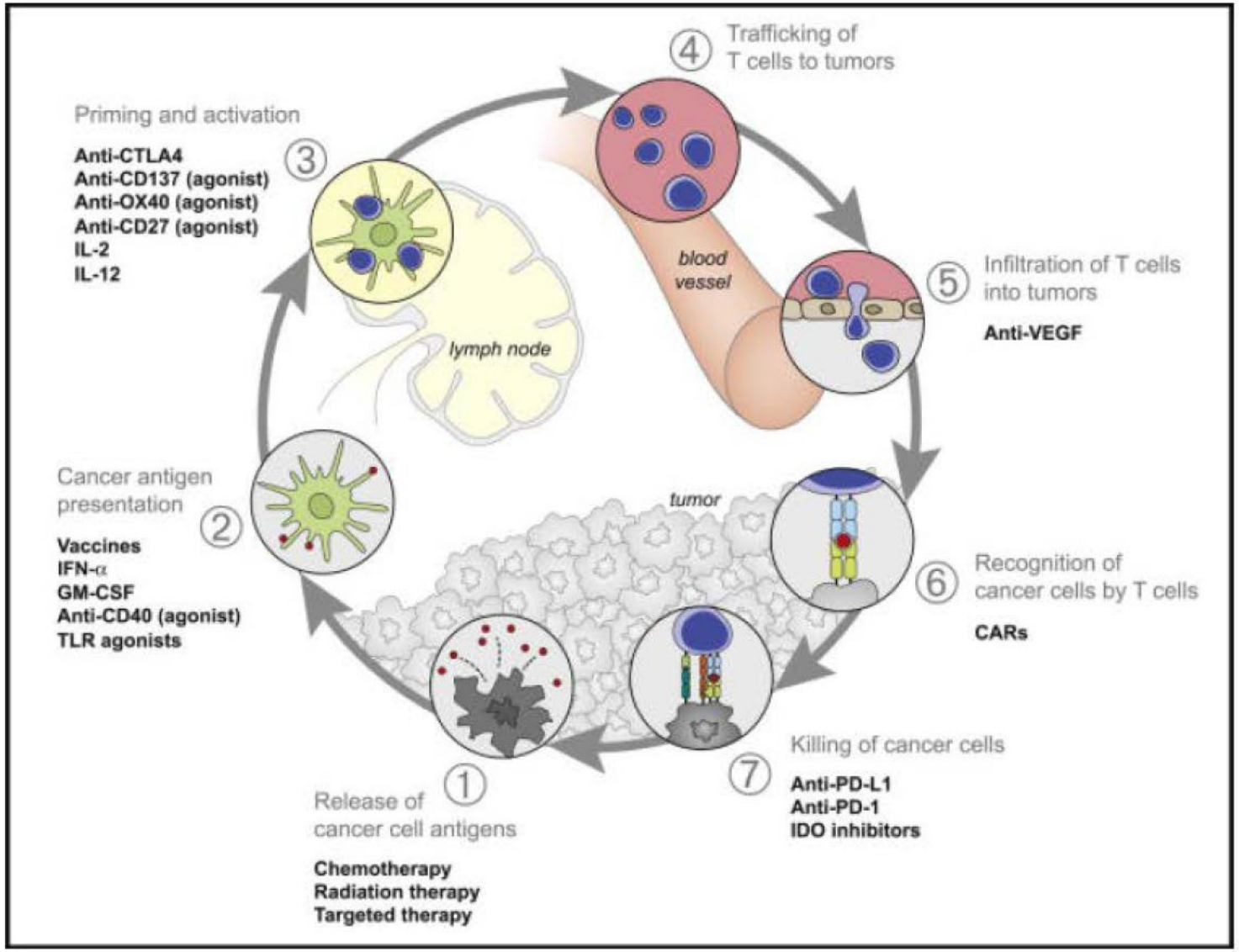
- Master protocols in "tumor agnostic" context
- Design operating characteristics of MEM models
- R package for Multisource Exchangeability Models

Traditional Clinical Research Paradigm

- average of 9.1 years regulatory approval between 1990 and 2005 [1]
- recent data (1998 - 2014) suggest that timelines for cytotoxic agents remained slow (median of 9.4 years) [2]
- foundational assumptions pertaining to **dose-response** and **inter-patient exchangeability**

1. DiMasi JA, Grabowski HG. Economics of new oncology drug development. *J Clin Oncol.* 2007

2. Jardim DL, Schwaederle M, Hong DS, Kurzrock R. An appraisal of drug development timelines in the Era of precision oncology. *Oncotarget.* 2016;7(33):53037-53046.

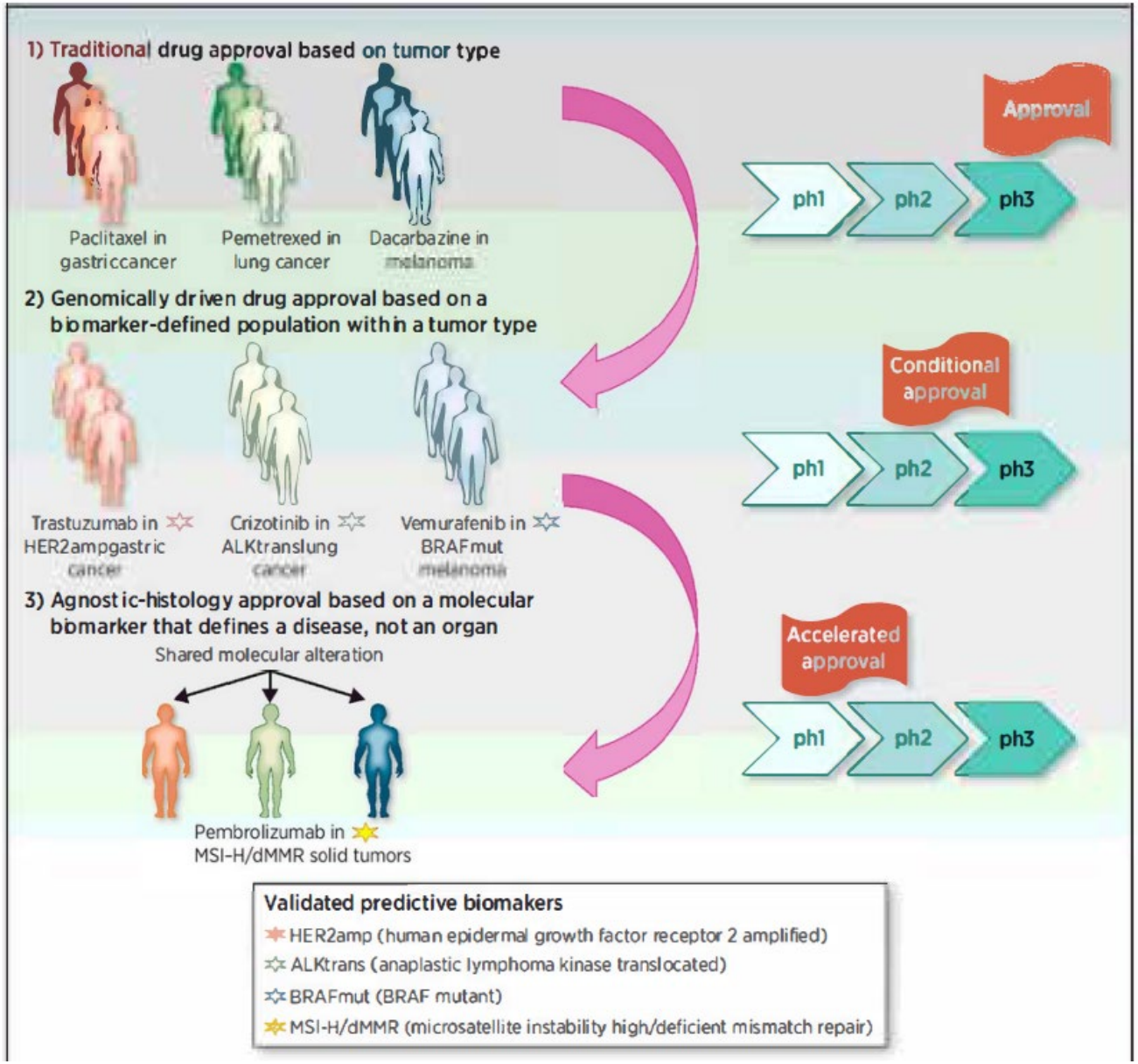


- Chen and Mellman (2013). "Oncology meets immunology: the cancer-immunity cycle"

Recent emphasis on non-cytotoxics

- partitions cancers into many small molecular subtypes
- drugs target molecular mechanism across many indications
- pathways mediating proliferation and cell survival or promote anti-cancer immunity
- Between 1998 – 2014 targeted therapies median duration of 5.4 yrs*
  - **pembrolizumab received accelerated approval in only 3.7 years ***

* Jardim DL, Schwaederle M, Hong DS, Kurzrock R. An appraisal of drug development timelines in the Era of precision oncology. *Oncotarget.* 2016;7(33):53037-53046.



CLINICAL CANCER RESEARCH

Home About Articles For Authors Alerts News Search

Review

Agnostic-Histology Approval of New Drugs in Oncology: Are We Already There?

Cinta Hierro, Ignacio Matos, Juan Martín-Liberal, María Ochoa de Olza, and Elena Garralda

Review Article | Published: 11 June 2020

Histology-agnostic drug development — considering issues beyond the tissue

Roberto Carmagnani Pestana, Shiraj Sen, Brian P. Hobbs & David S. Hong

Nature Reviews Clinical Oncology **17**, 555–568(2020)

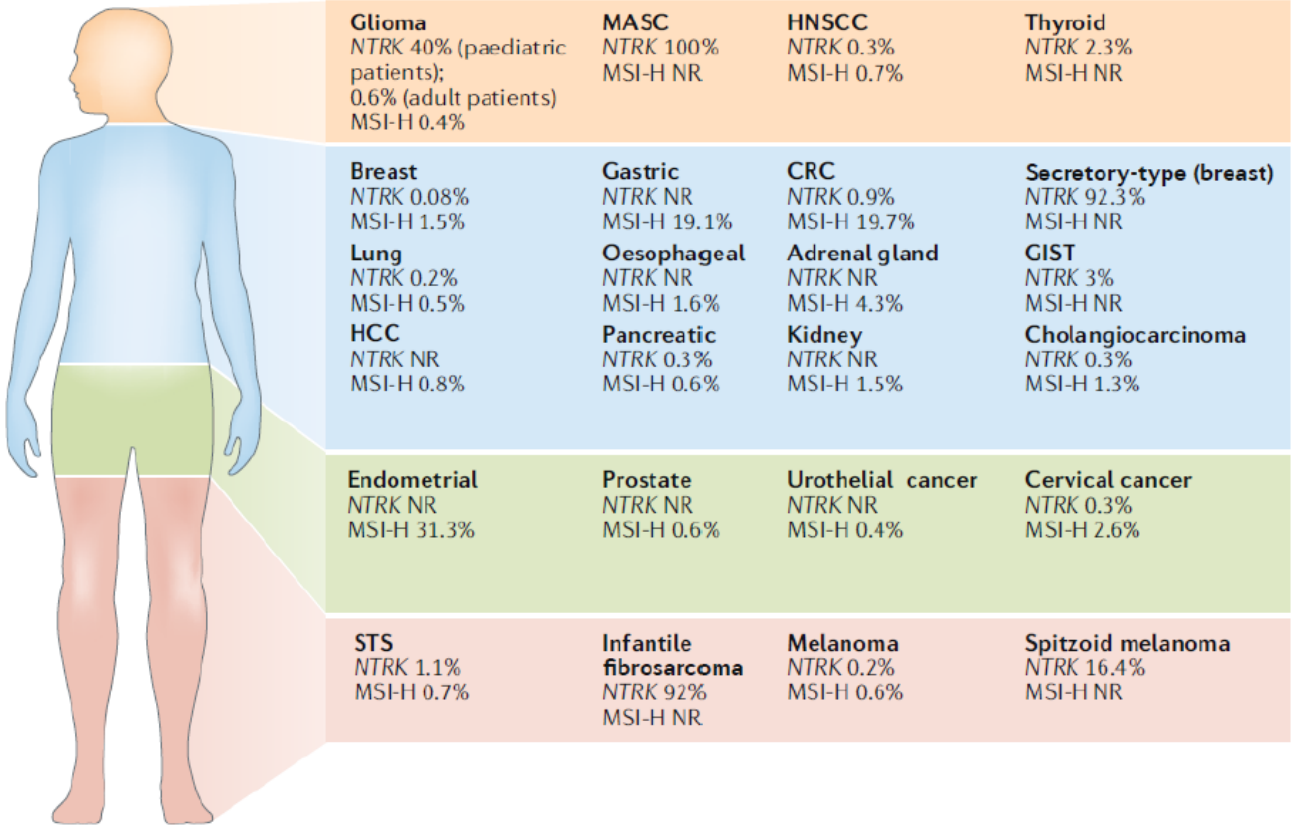


Fig. 2 | **Prevalence of specific alterations for which histology-agnostic drugs are approved across tumour types**[33,57,125,153–156]. *NTRK* gene fusions and microsatellite instability-high (MSI-H) or mismatch-repair deficiency (dMMR) phenotypes are present across multiple tumour types. Knowledge of the prevalence of these features in each tumour

Challenges for non-cytotoxic development

I. Development for individual histologies may be **prohibitive**

II. Requires **reproducible** diagnostic/biomarker assay

III. Can one molecular label describe an "**exchangeable**" patient cohort?

- tumor initiating cells derive from tissue-specific stem cells
- proliferation rates of tumor cells vary by histology

IV. Therapies may lack conventional/**monotonic a dose-response**

FDA Officials: Master Protocols Needed for Precision Medicine

Posted 07 July 2017 | By Michael Mezher

In an article published Thursday in the *New England Journal of Medicine*, two top officials from the US Food and Drug Administration (FDA) say that "master protocols" for studies involving multiple drugs or multiple diseases (or both) simultaneously are needed to efficiently generate evidence for precision medicines.

"The standard approach to generating this evidence—a series of clinical trials, each investigating one or two interventions in a single disease—has become ever more expensive and challenging to execute. As a result, important clinical questions go unanswered," write Janet Woodcock, director of the Center for Drug Evaluation and Research (CDER) and Lisa LaVange, director of Office of Biostatistics within CDER.

Instead, the two argue that well-designed master protocols that look at multiple therapies in a single disease, a single therapy in multiple diseases, or multiple therapies across multiple diseases or disease subtypes, can provide answers more quickly and efficiently than traditional "stand-alone" clinical trials.

The NEW ENGLAND JOURNAL of MEDICINE

REVIEW ARTICLE

THE CHANGING FACE OF CLINICAL TRIALS

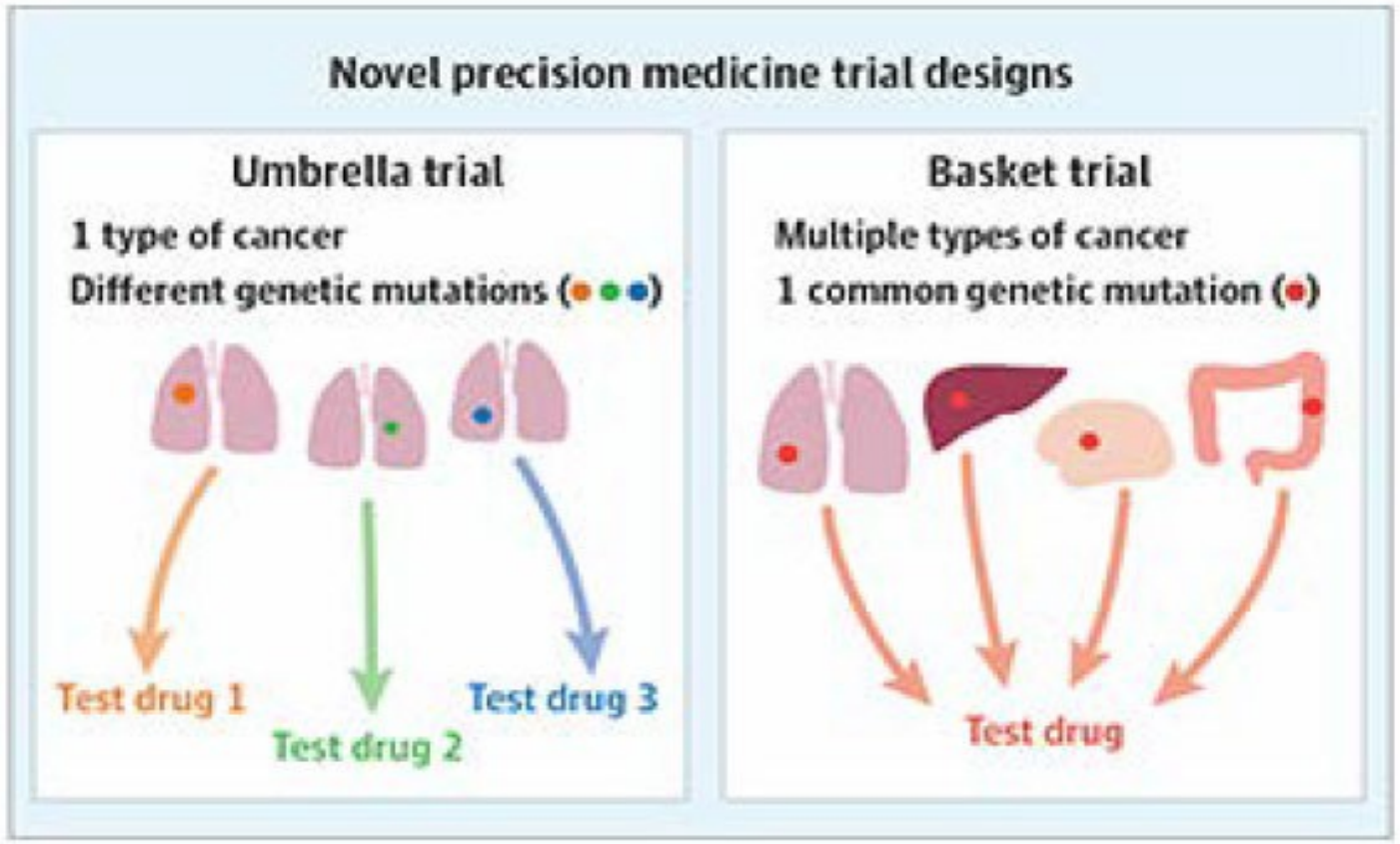
Jeffrey M. Drazen, M.D., David P. Harrington, Ph.D., John J.V. McMurray, M.D., James H. Ware, Ph.D., and Janet Woodcock, M.D., *Editors*

Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both

Janet Woodcock, M.D., and Lisa M. LaVange, Ph.D.

HIGH-QUALITY EVIDENCE IS WHAT WE USE TO GUIDE MEDICAL PRACTICE. The standard approach to generating this evidence — a series of clinical trials, each investigating one or two interventions in a single disease — has become ever more expensive and challenging to execute. As a result, important clinical questions go unanswered. The conduct of "precision medicine" trials to evaluate targeted therapies creates challenges in recruiting patients with rare genetic subtypes of a disease. There is also increasing interest in performing mechanism-based trials in which eligibility is based on criteria other than traditional disease definitions. The common denominator is a need to answer more questions more efficiently and in less time.

A methodologic innovation responsive to this need involves coordinated efforts to evaluate more than one or two treatments in more than one patient type or disease within the same overall trial structure.[1-4] Such efforts are referred to as master protocols, defined as one overarching protocol designed to answer multiple questions. Master protocols may involve one or more interventions in multiple diseases or a single disease, as defined by current disease classification, with multiple interventions, each targeting a particular biomarker-defined population or disease subtype. Included under this broad definition of a master protocol are three distinct entities: umbrella, basket, and platform trials (Table 1 and Figs. 1 and 2). All constitute a collection of trials or substudies that share key design components and operational aspects to achieve better coordination than can be achieved in single trials designed



JAMA Oncology: doi:10.1001/jamaoncol.2016.5299

multiple diseases or disease subtypes, can provide answers more quickly and efficiently than traditional "stand-alone" clinical trials.

The NEW ENGLAND JOURNAL of MEDICINE

REVIEW ARTICLE

THE CHANGING FACE OF CLINICAL TRIALS

Jeffrey M. Drazen, M.D., David P. Harrington, Ph.D., John J.V. McMurray, M.D., James H. Ware, Ph.D., and Janet Woodcock, M.D., *Editors*

Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both

Janet Woodcock, M.D., and Lisa M. LaVange, Ph.D.

HIGH-QUALITY EVIDENCE IS WHAT WE USE TO GUIDE MEDICAL PRACTICE. The standard approach to generating this evidence — a series of clinical trials, each investigating one or two interventions in a single disease — s become ever more expensive and challenging to execute. As a result, important nical questions go unanswered. The conduct of "precision medicine" trials to evalu- targeted therapies creates challenges in recruiting patients with rare genetic btypes of a disease. There is also increasing interest in performing mechanism-sed trials in which eligibility is based on criteria other than traditional disease finitions. The common denominator is a need to answer more questions more ef-iently and in less time.

A methodologic innovation responsive to this need involves coordinated efforts evaluate more than one or two treatments in more than one patient type or disease thin the same overall trial structure.[1-4] Such efforts are referred to as master pro-cols, defined as one overarching protocol designed to answer multiple questions. Master protocols may involve one or more interventions in multiple diseases or a single disease, as defined by current disease classification, with multiple interventions, each targeting a particular biomarker-defined population or disease subtype. In-cluded under this broad definition of a master protocol are three distinct entities: umbrella, basket, and platform trials (Table 1 and Figs. 1 and 2). All constitute a collection of trials or substudies that share key design components and operational aspects to achieve better coordination than can be achieved in single trials designed

Who Can be Averaged?

Precision Medicine from the perspective of Data = ascertaining statistical exchangeability

Received: 11 August 2017 | Revised: 17 March 2018 | Accepted: 8 June 2018
DOI: 10.1002/sim.7893

RESEARCH ARTICLE

WILEY Statistics in Medicine

Bayesian basket trial design with exchangeability monitoring

Brian P. Hobbs[1] | Rick Landin[2]

[1]Department of Quantitative Health Sciences and the Taussig Cancer Institute, Cleveland Clinic, Cleveland, Ohio 44195

[2]La Jolla Pharmaceutical Company, San Diego, California 92121

Correspondence
Brian P. Hobbs, Department of Quantitative Health Sciences and The Taussig Cancer Institute, Cleveland Clinic, Cleveland, OH 44195.
Email: BPHobbs@gmail.com

Precision medicine endeavors to conform therapeutic interventions to the individuals being treated. Implicit to the concept of precision medicine is heterogeneity of treatment benefit among patients and patient subpopulations. Thus, precision medicine challenges conventional paradigms of clinical translational which have relied on estimates of population-averaged effects to guide clinical practice. Basket trials comprise a class of experimental designs used to study solid malignancies that are devised to evaluate the effectiveness of a therapeutic strategy among patients defined by the presence of a particular drug target (often a genetic mutation) rather than a particular tumor histology. Acknowledging the potential for differential effectiveness on the basis of traditional criteria for cancer subtyping, evaluations of treatment effectiveness are con-

The R Journal

Analyzing Basket Trials under Multisource Exchangeability Assumptions

Michael J. Kane
Yale University

Nan Chen
The University of Texas

Alexander M. Kaizer
University of Colorado

Xun Jiang
Amgen Inc.

H. Amy Xia
Amgen Inc.

Brian P. Hobbs
Cleveland Clinic

Abstract

Basket designs are prospective clinical trials that are devised with the hypothesis that the presence of selected molecular features determine a patient's subsequent response to a particular "targeted" treatment strategy. Basket trials are designed to enroll multiple clinical subpopulations to which it is assumed that the therapy in question offers beneficial efficacy in the presence of the targeted molecular profile. The treatment, however, may not offer acceptable efficacy to all subpopulations enrolled. Moreover, for rare dis-

Case Study: Vemurafenib non-melanoma basket trial

Baskets	Enrolled	Evaluable	Responders	Posterior probability $Pr(\pi > 0.15)$ based on response only
NSCLC	20	19	8	**0.998**
CRC (vemu)	10	10	0	**0.068**
CRC (vemu + cetu)	27	26	1	**0.039**
Bile Duct	8	8	1	**0.472**
ECD or LCH	18	14	6	**0.995**
ATC	7	7	2	**0.847**

Histology-independent phase 2 basket trial of vemurafenib in BRAF V600 mutation-positive non-melanoma cancers

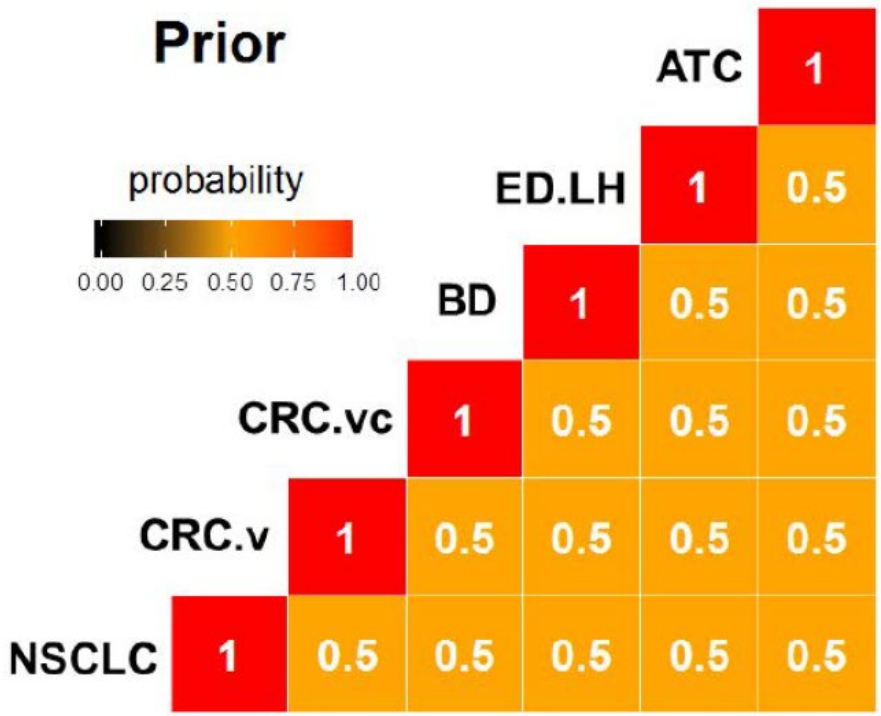
Six pre-specified cancer cohorts

Adaptive Simon two-stage design with primary endpoint of response rate at week 8

Basket-wise testing: $H_0 : \pi = 0.15$ vs. $H_1 : \pi = 0.35$

Bayesian Basket Trial Design with Exchangeability Monitoring

Brian P. Hobbs[1] **and Rick Landin**[2]



Case Study Analysis: Vemurafenib non-melanoma basket trial

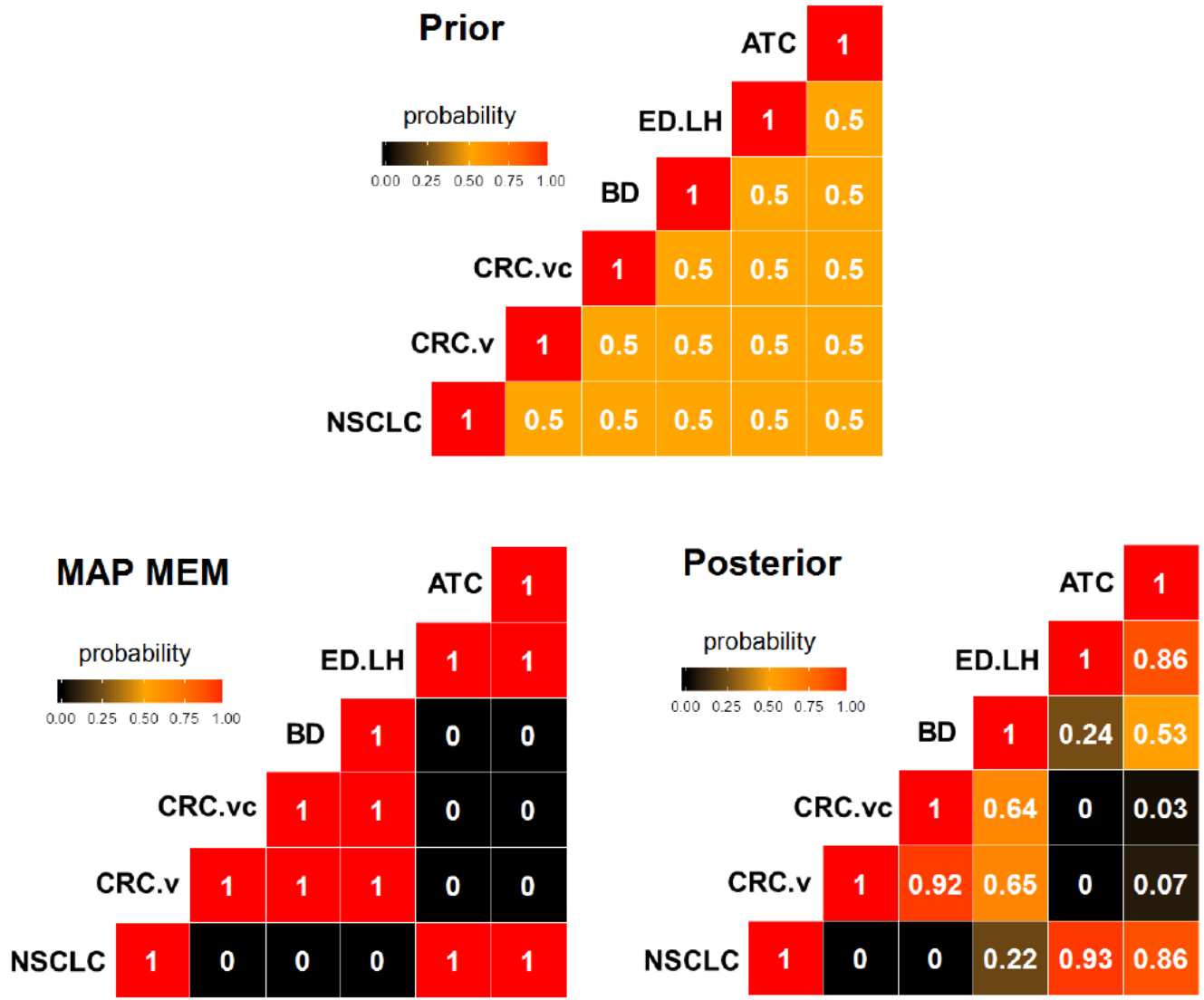


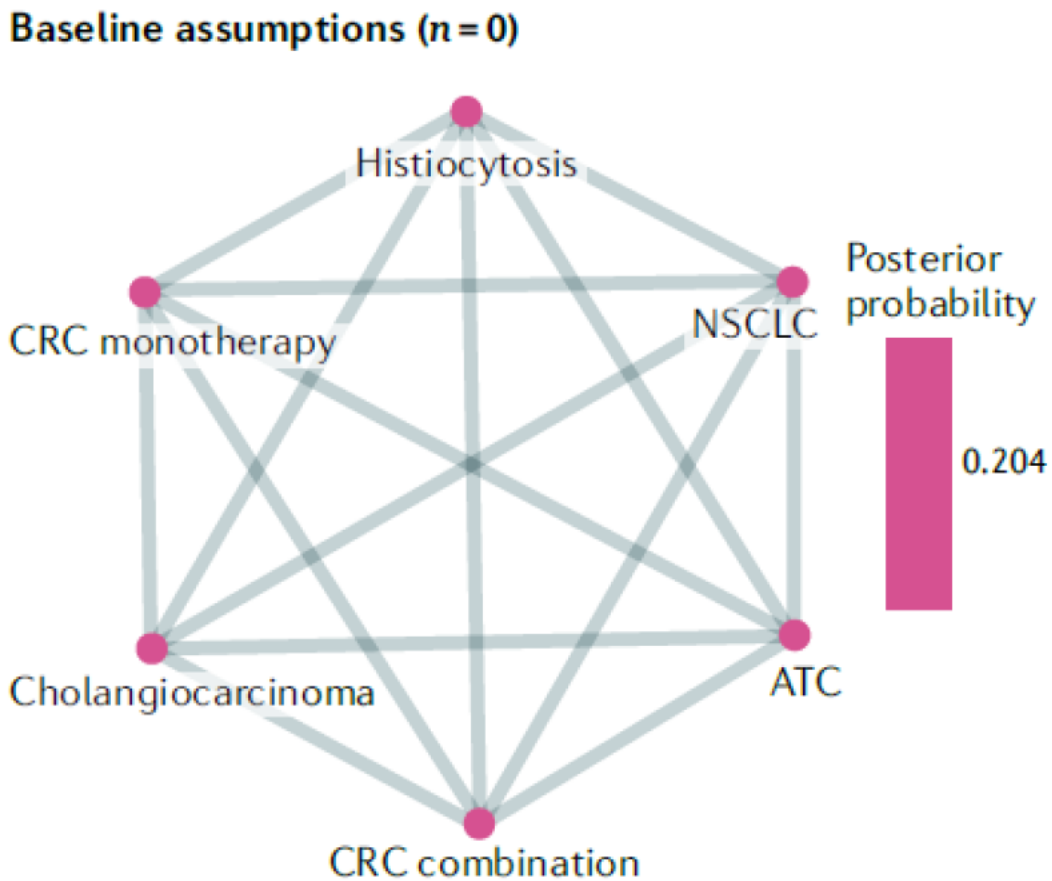
Figure 2. Prior, MAP, and PEP that result from Bayesian inference using the observed vemurafenib basket trial data

Review Article | Published: 11 June 2020

Histology-agnostic drug development — considering issues beyond the tissue

Roberto Carmagnani Pestana, Shiraj Sen, Brian P. Hobbs & David S. Hong

Nature Reviews Clinical Oncology **17**, 555–568(2020)

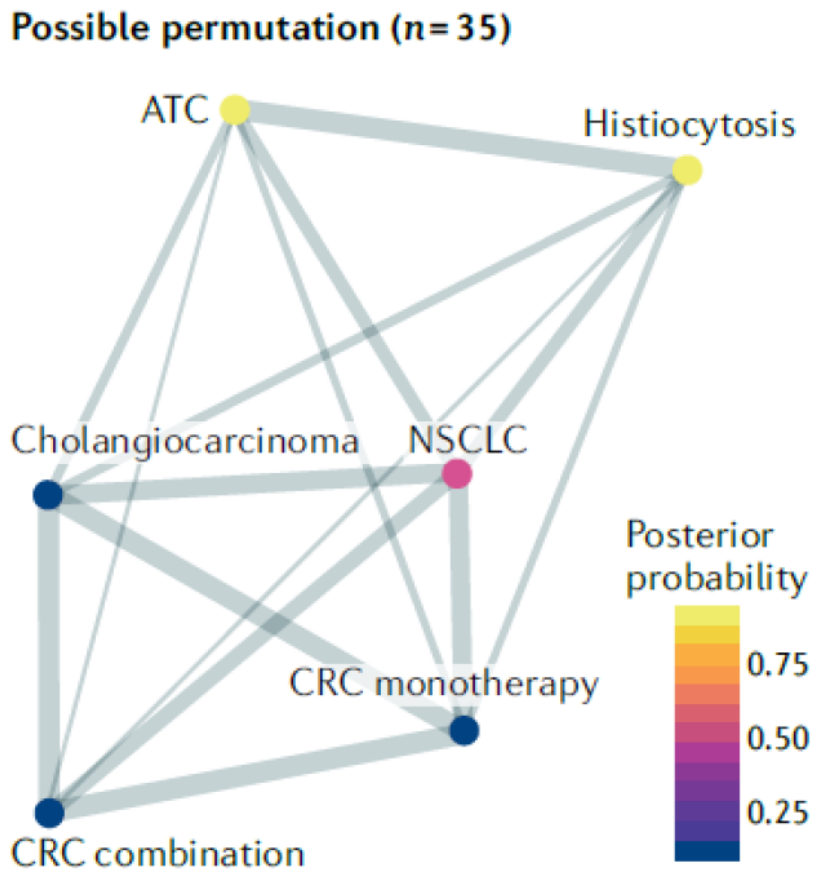


Review Article | Published: 11 June 2020

Histology-agnostic drug development — considering issues beyond the tissue

Roberto Carmagnani Pestana, Shiraj Sen, Brian P. Hobbs & David S. Hong

Nature Reviews Clinical Oncology **17**, 555–568(2020)

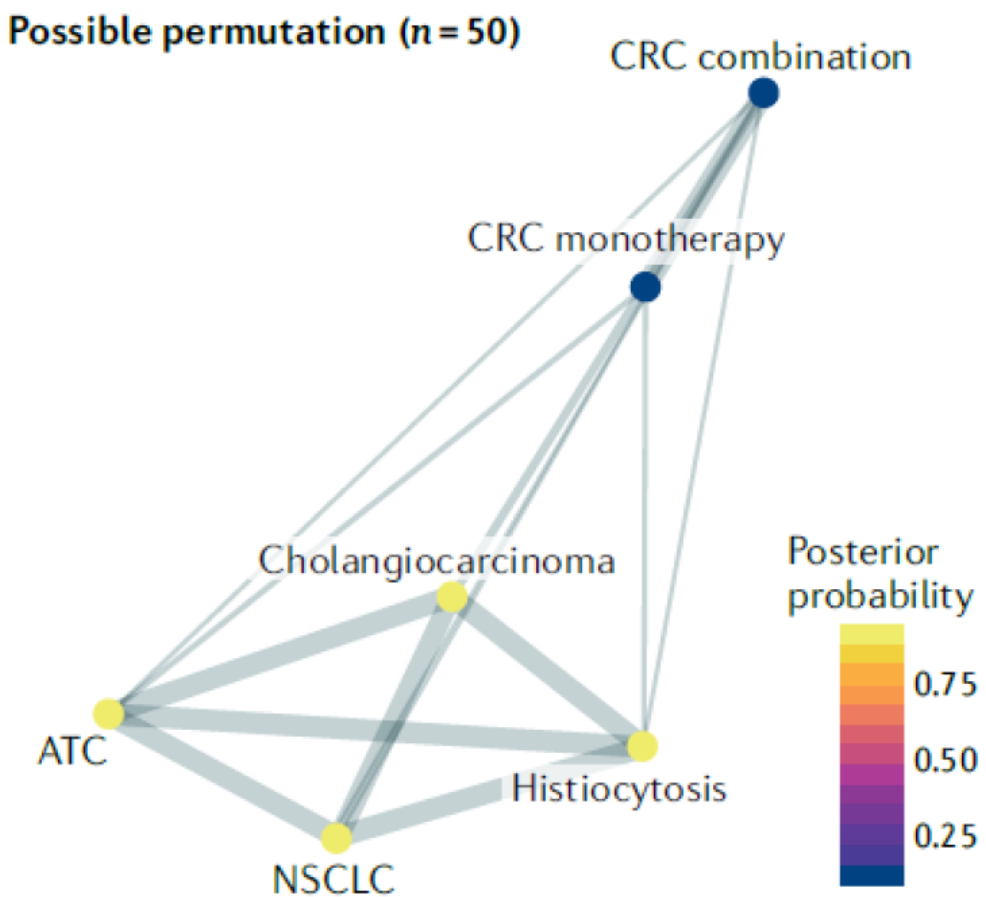


Review Article | Published: 11 June 2020

Histology-agnostic drug development — considering issues beyond the tissue

Roberto Carmagnani Pestana, Shiraj Sen, Brian P. Hobbs & David S. Hong

Nature Reviews Clinical Oncology **17**, 555–568(2020)

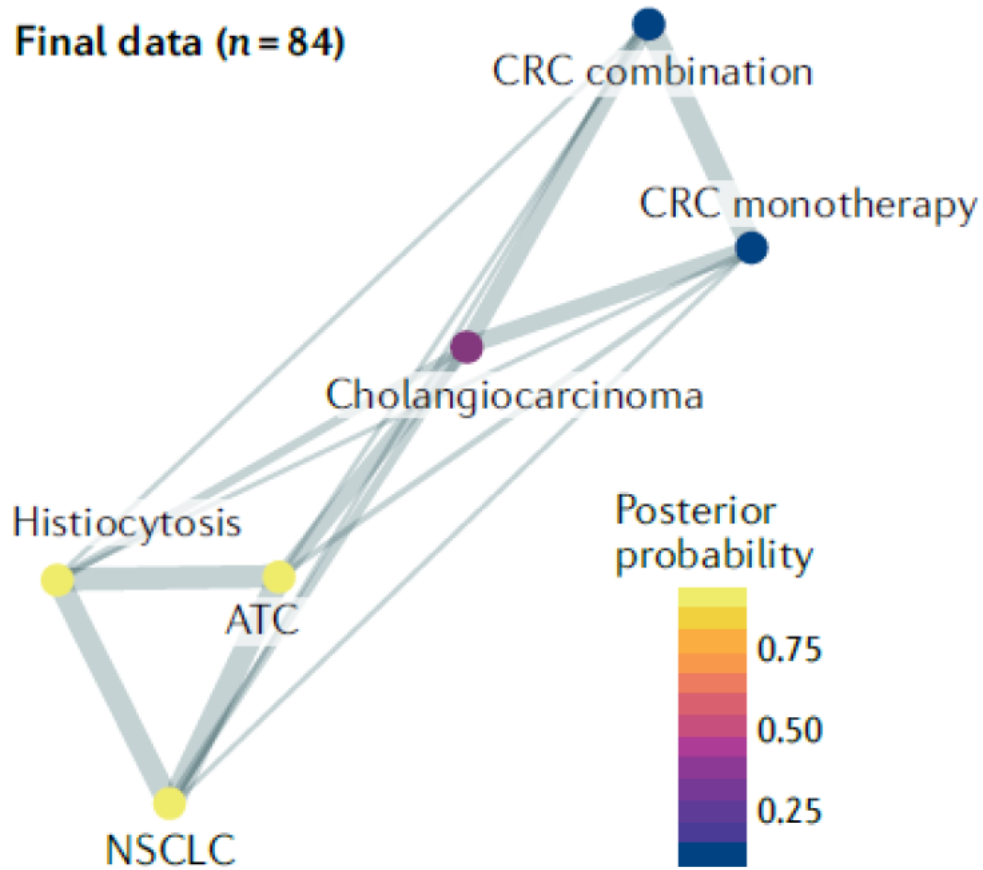


Review Article | Published: 11 June 2020

Histology-agnostic drug development — considering issues beyond the tissue

Roberto Carmagnani Pestana, Shiraj Sen, Brian P. Hobbs & David S. Hong

Nature Reviews Clinical Oncology **17**, 555–568(2020)



2. Exchangeability for Trials with Subpopulations

2.1. The Single-Source Exchangeability Model

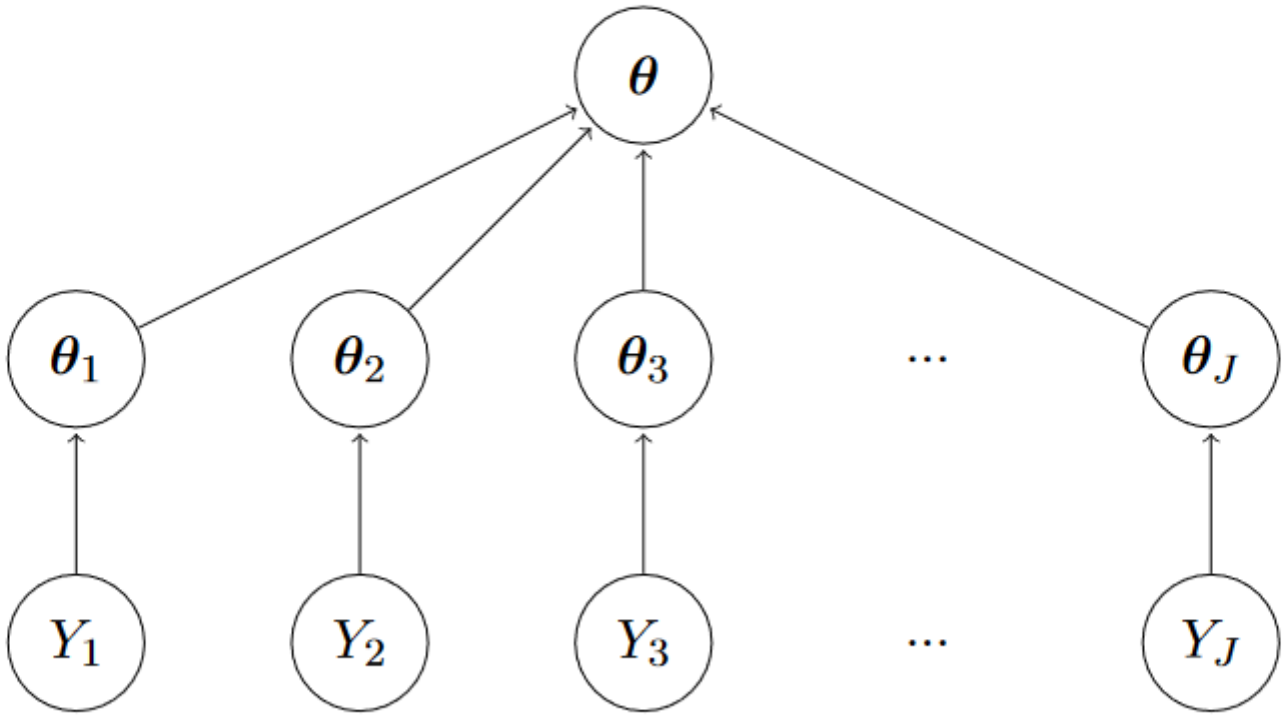
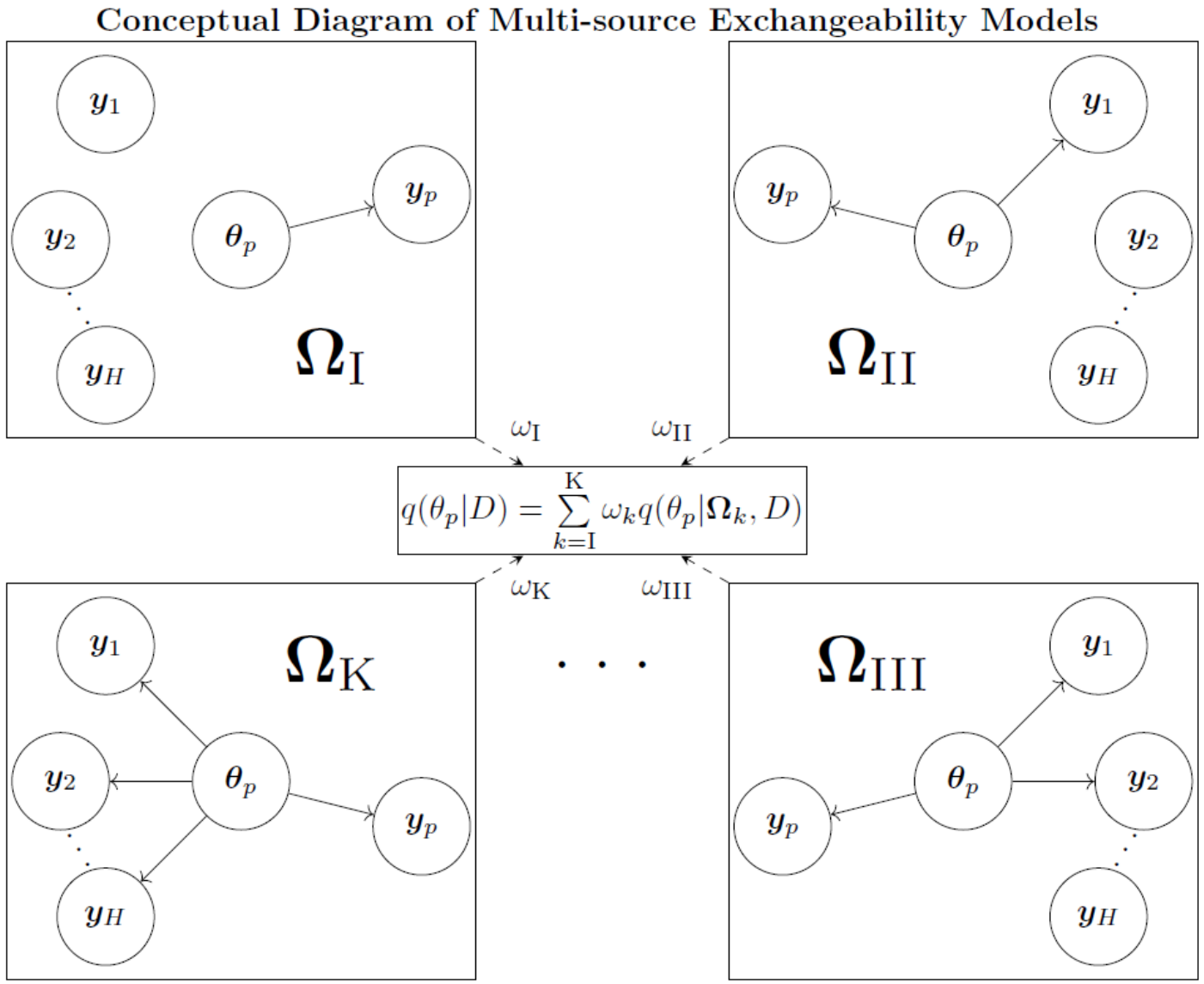


Figure 1: A conventional single-source Bayesian hierarchical model with J subtypes.

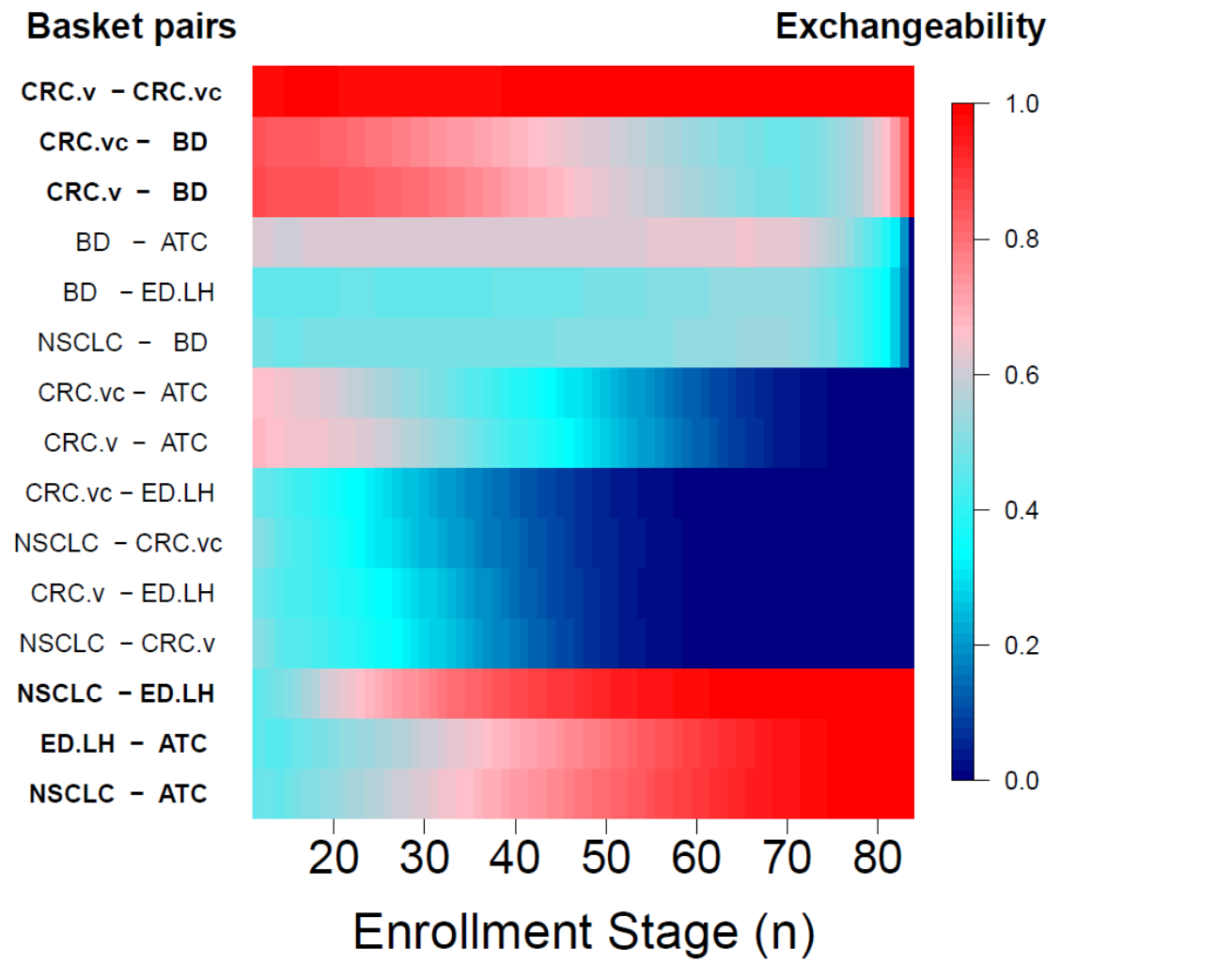
Multisource Exchangeability Model



Multisource Exchangeability Models

- Asymmetric settings (primary & supple cohorts); Kaizer et al. 2017, Biostatistics
- Symmetric settings (all cohorts primary) and sequential design; Hobbs and Landin 2018, Stat in Med
- Adaptive Platform Design; Kaizer et al. 2018, Biometrics
- Frequentist Trial Operating Characteristics; Kaizer et al. 2019, JCO Precision Oncology
- Open-source statistical software with the Basket package; Kane et al. 2020, The R Journal
- Design optimization for trials with multiple indications; Kaizer et al. 2020, SMMR

Sequential Design with MEM



The Challenge of

Patient Heterogeneity

Confounds statistical inference of drug-target effectiveness across diverse subpopulations

ANNALS OF ONCOLOGY

ESMO GOOD SCIENCE BETTER MEDICINE BEST PRACTICE

Article Contents

Abstract

ACCEPTED MANUSCRIPT

Statistical challenges posed by uncontrolled master protocols: sensitivity analysis of the Vemurafenib study FREE

B P Hobbs, M J Kane, D S Hong, R Landin

Annals of Oncology, mdy457, https://doi.org/10.1093/annonc/mdy457

Published: 18 October 2018

Abstract

Within the evidentiary hierarchy of experimental inquiry, randomized trials are the gold standard. Oncology patients enter clinical studies with diverse lifestyles, treatment pathways, host tissue environments, and competing co-morbidities. Randomization attempts to balance prognostic characteristics among study arms, thereby enabling statistical inference of "average benefit" and attribution to the studied therapies. In contrast, interpretations of uncontrolled trials require additional scrutiny to attempt to place the findings in the context of external evidence. Counter-factual reasoning and speculation across trials may be obscured by the disproportionate enrollment of prognostic subpopulations which may be unknown from publications of trial reports. Recent modifications to the regulatory environment (Food and Drug Administration Safety and Innovation

Acknowledgements

Trainees

Caimiao Wei (Pfizer),

Shabnam Azadeh (FDA),

Meilin Huang (Genetech),

Xiao Li (Genetech),

Yuan Wang (Assist Prof Washington St.)

Junsheng Ma (MD Anderson)

Jiera Zhou (Cleveland Clinic)

Collaborators

David Hong (MD Anderson)

Chad Tang (MD Anderson)

Nan Chen (Gilead)

Joe Koopmeiners(Minnesota)

Alex Kaizer (Colorado)

Michael Kane (Yale)

Emily Zabor (Cleveland Clinic)

Multisource Exchangeability Models

hyperparameters a and b, upon having observed successes $\boldsymbol{S} = \{S_1, \ldots, S_J\}$, Bayes' Theorem yields the following conjugate conditional posterior distribution for the response probability of basket j representing the Bayesian update of $p(\boldsymbol{\pi} \mid \boldsymbol{S}_{(-j)})$ with likelihood $Bin(S_j \mid \pi_j, n_j)$,

$$q(\pi_j|\boldsymbol{S}, \boldsymbol{\Omega}_j) \propto \text{Beta}\left(a + \sum_{h=1}^{J} \Omega_{j,h} S_h,\ b + \sum_{k=1}^{J} \Omega_{j,k}(n_k - S_k)\right), \tag{1}$$

where $\boldsymbol{\Omega}_j$ represents the j^{th} row of multisource exchangeability matrix $\boldsymbol{\Omega}$. Marginal posterior inference with respect to $\pi_j \mid \boldsymbol{S}$ averages (1) with respect to the marginal posterior probability of $G = 2^{J-1}$ possible exchangeability configurations of $\boldsymbol{\Omega}_j$. Let $\boldsymbol{\omega} = \{\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_G\}$ denote the collection of vectors each of length J and with j^{th} element $= 1$ that collectively span the sample space of $\boldsymbol{\Omega}_j$. The marginal posterior distribution can be represented by a finite mixture density

$$q(\pi_j|\boldsymbol{S}) \propto \sum_{g=1}^{G} q(\pi_j \mid \boldsymbol{S}, \boldsymbol{\Omega}_j = \boldsymbol{\omega}_g) Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_g \mid \boldsymbol{S}), \tag{2}$$

where the posterior probability of exchangeability configuration $\boldsymbol{\omega}_g$ given the observed data follows from Bayes' Theorem in proportion to the marginal density of the data given $\boldsymbol{\omega}_g$ and its unconditional prior probability

$$Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_g \mid \boldsymbol{S}) \propto \frac{m(\boldsymbol{S}_j \mid \boldsymbol{\Omega}_j = \boldsymbol{\omega}_g,\ \boldsymbol{S}_{(-j)}) Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_g)}{\sum_{u=1}^{G} m(\boldsymbol{S}_j \mid \boldsymbol{\Omega}_j = \boldsymbol{\omega}_u,\ \boldsymbol{S}_{(-j)}) Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_u)}. \tag{3}$$

Multisource Exchangeability Models

hyperparameters a and b, upon having observed successes $\boldsymbol{S} = \{S_1, \ldots, S_J\}$, Bayes' Theorem yields the following conjugate conditional posterior distribution for the response probability of basket j representing the Bayesian update of $p(\boldsymbol{\pi} \mid \boldsymbol{S}_{(-j)})$ with likelihood $Bin(S_j \mid \pi_j, n_j)$,

$$q(\pi_j|\boldsymbol{S}, \boldsymbol{\Omega}_j) \propto \text{Beta}\left(a + \sum_{h=1}^{J} \Omega_{j,h} S_h,\ b + \sum_{k=1}^{J} \Omega_{j,k}(n_k - S_k)\right), \tag{1}$$

where $\boldsymbol{\Omega}_j$ represents the j^{th} row of multisource exchangeability matrix $\boldsymbol{\Omega}$. Marginal posterior inference with respect to $\pi_j \mid \boldsymbol{S}$ averages (1) with respect to the marginal posterior probability of $G = 2^{J-1}$ possible exchangeability configurations of $\boldsymbol{\Omega}_j$. Let $\boldsymbol{\omega} = \{\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_G\}$ denote the collection of vectors each of length J and with j^{th} element $= 1$ that collectively span the sample space of $\boldsymbol{\Omega}_j$. The marginal posterior distribution can be represented by a finite mixture density

$$q(\pi_j|\boldsymbol{S}) \propto \sum_{g=1}^{G} q(\pi_j \mid \boldsymbol{S}, \boldsymbol{\Omega}_j = \boldsymbol{\omega}_g) Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_g \mid \boldsymbol{S}), \tag{2}$$

where the posterior probability of exchangeability configuration $\boldsymbol{\omega}_g$ given the observed data follows from Bayes' Theorem in proportion to the marginal density of the data given $\boldsymbol{\omega}_g$ and its unconditional prior probability

$$Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_g \mid \boldsymbol{S}) \propto \frac{m(\boldsymbol{S}_j \mid \boldsymbol{\Omega}_j = \boldsymbol{\omega}_g,\ \boldsymbol{S}_{(-j)}) Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_g)}{\sum_{u=1}^{G} m(\boldsymbol{S}_j \mid \boldsymbol{\Omega}_j = \boldsymbol{\omega}_u,\ \boldsymbol{S}_{(-j)}) Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_u)}. \tag{3}$$

Multisource Exchangeability Models

Let $B()$ denote the beta function. Given an exchangeability configuration $\boldsymbol{\Omega}_j$, the marginal density of $\boldsymbol{S}_j$ may be obtained by integrating the likelihood of $\pi_j \mid S_j$ with respect to $p(\boldsymbol{\pi} \mid \boldsymbol{S}_{(-j)})$,

$$m(\boldsymbol{S}_j \mid \boldsymbol{\Omega}_j,\ \boldsymbol{S}_{(-j)}) \propto \frac{B\left(a + \sum_{h=1}^{J} \Omega_{j,h} S_h,\ b + \sum_{k=1}^{J} \Omega_{j,k}(n_k - S_k)\right)}{B(a,b)} \times \prod_{i=1}^{J} \left(\frac{B(a+S_i,\ b+n_i-S_i)}{B(a,b)}\right)^{1-\Omega_{j,i}}. \tag{4}$$

The Bayesian model is complete given specification of a vector comprising the unconditional prior probabilities of all possible pairwise exchangeability configurations, $Pr(\boldsymbol{\Omega})$, which is challenging given the high dimensionality of the MEM sample domain. Defining $Pr(\boldsymbol{\Omega})$ as the product of prior exchangeability probabilities for each unique basket pair, however, reduces the dimension from $\prod_{k=1}^{J-1} 2^k$ to $J(J-1)/2$ yielding feasibility and thereby offering an advantage with respect to conventional Bayesian model averaging

$$Pr(\boldsymbol{\Omega}) = Pr(\Omega_{1,2} = 1) \times Pr(\Omega_{1,3} = 1) \times \cdots \times Pr(\Omega_{J-1,J} = 1). \tag{5}$$

By the Kolmogorov definition of conditional probability, the prior exchangeability probabilities for all 2^{J-1} configurations of $\boldsymbol{\Omega}_j$ in (3) follow from (5) as $Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}) = \prod_{i=1}^{J} Pr(\Omega_{j,i} = 1)^{I(\omega_i=1)} \times \{1 - Pr(\Omega_{j,i} = 1)\}^{(1-I(\omega_i=1))}$, where $I()$ is the indicator function and $\boldsymbol{\omega}$ represents one vector of length J within the sample domain of $\boldsymbol{\Omega}_j$.

Multisource Exchangeability Models

2.3. Posterior Probability and Effective Sample Size

The MEM Bayesian model specification facilitates posterior inference with respect to all possible pairwise exchangeability relationships among J subtypes. The framework facilitates estimation of disjointed subpopulations comprised of meta-subtypes or singelton subtypes and thereby offers additional flexibility when compared to SEM specifications. The uncontrolled basket study considered herein is devised with the intention of testing the hypothesis that the response probability for a targeted intervention exceeds a null value, which we denote π_0, while acknowledging the potential for heterogeneity in effectiveness in accordance with the pre-specified basket partitions. Within the MEM framework, this testing procedure follows from the cumulative density function (cdf) of the marginal posterior distribution (2). Specifically, the posterior probability that π_j exceeds π_0 may be computed as the weighted average of cdfs for all possible exchangeability configurations,

$$Pr(\pi_j > \pi_0 \mid \boldsymbol{S}) = \sum_{g=1}^{G} Pr(\boldsymbol{\Omega}_j = \boldsymbol{\omega}_g \mid \boldsymbol{S}) \left\{ 1 - \frac{\int_0^{\pi_0} u^{a+\sum_{h=1}^{J} \omega_{g,h} S_h - 1} (1-u)^{b+\sum_{k=1}^{J} \omega_{g,k}(n_k - S_k) - 1} du}{B(a + \sum_{h=1}^{J} \omega_{g,h} S_h,\ b + \sum_{k=1}^{J} \omega_{g,k}(n_k - S_k) - 1)} \right\}. \qquad (6)$$

Basket Trial Design: Challenges and Potential Solutions

Alex Kaizer

Assistant Professor
Department of Biostatistics and Informatics
Colorado School of Public Health

DIA BSWG

Graduate School
UNIVERSITY OF COLORADO
DENVER | ANSCHUTZ MEDICAL CAMPUS

Discussed Papers

- Kaizer AM, Koopmeiners JS, Kane MJ, Roychoudhury S, Hong DS, Hobbs BP. “Basket designs: Statistical considerations for oncology trials.” *JCO Precision Oncology*, 2019.
- Kaizer AM, Koopmeiners JS, Chen N, Hobbs BP. “Statistical design considerations for trials that study multiple indications.” *Statistical Methods in Medical Research*, 2020.
- Kaizer AM, Koopmeiners JS, Hobbs BP. “Bayesian hierarchical modeling based on multisource exchangeability.” *Biostatistics*, 2018.

Motivation

- Trial design has evolved alongside the scientific developments leading to precision medicine in oncology
- Master protocols provide a flexible approach to designs with multiple indications, but have their own challenges (Woodcock and LaVange, 2017; Hobbs et al., 2018; Kaizer et al., 2019)
- Examples of master protocols include basket, umbrella, and platform trials

Challenges

- Often we have small *n* for each indication (e.g., basket)
- Indication/subgroup heterogeneity is likely (i.e., we might not expect the same response in all baskets)
- Operating characteristics, such as power and type I error rates, have special considerations
- Designing multi-indication studies generally requires optimizing for one scenario

Motivating Example: Basket Trial with 5 Baskets

Scenarios	Basket Number				
	1	2	3	4	5
Global Null 1					
2					
3					
4					
5					
6 *Global Alternative*					

LEGEND: Null Basket Alternative Basket

Statistical Considerations

Basket Number				
1	2	3	4	5
LEGEND: Null			Alternative	

Subgroup analysis:

- Pooled (combine all 5 baskets)
- Independent (analyze each separately)
- Information sharing between baskets (exchangeability) (Hobbs and Landin, 2018)

Statistical Considerations

Basket Number				
1	2	3	4	5
LEGEND: Null			Alternative	

Subgroup analysis (pooled, independent, exchangeable)

Type I error rate:

- Marginal (basket-specific outcomes)
- Family-wise (considering all basket outcomes)

Statistical Considerations

Basket Number				
1	2	3	4	5
LEGEND: Null			Alternative	

Subgroup analysis (pooled, independent, exchangeable)

Type I error rate (marginal vs. family-wise)

Strength of type I error control (Dmitrienko et al., 2009):

- Weak (family-wise type I error for global null)
- Strong (family-wise type I error for any scenario)

How to choose the optimal approach?

- Optimal model for a basket trial depends upon the context

How to choose the optimal approach?

- Optimal model for a basket trial depends upon the context
- Often we pick one scenario in the design stage, but we could encounter any scenario in the future trial

How to choose the optimal approach?

- Optimal model for a basket trial depends upon the context
- Often we pick one scenario in the design stage, but we could encounter any scenario in the future trial
- **Challenge 1:** how to identify the optimal subgroup analysis approach
- **Challenge 2:** how to identify the optimal hyperparameter values given uncertainty of scenarios

Example Application to 5 Basket Trial Simulations

- Simulate 10,000 fixed-sample trials with 5 baskets each having $n = 25$ for each scenario in R
- Assume null response of 10% and alternative of 30%
- Compare 4 approaches:
  1. Independence with no correction for multiple testing ($\alpha = 0.10$, Ind)
  2. Independence with Bonferroni correction ($\alpha = \frac{0.10}{5} = 0.02$, Mult)
  3. Pooling all baskets ($\alpha = 0.10$, Pool)
  4. Bayesian approach that facilitates information sharing with multi-source exchangeability models (MEM) with UB=0.1 hyperparameter (Hobbs and Landin, 2018)
- For MEMs, explore posterior probability threshold calibrated to achieve different targets

Results with MEM PP Threshold of 86.25%

With a posterior probability (PP) threshold of **86.25%** for the MEM approach, we observe:

# Active Baskets	Ind		Mult	Pool	MEM
	FT	P			
0	40	-			
1	33	91			
2	26	91			
3	18	91			
4	9	91			
5	-	91			

Legend:
FT = Family-wise Type I Error Rate (%)
MT = Marginal Type I Error Rate (%)
P = Power (%)

Results with MEM PP Threshold of 86.25%

With a posterior probability (PP) threshold of **86.25%** for the MEM approach, we observe:

# Active Baskets	Ind		Mult		Pool	MEM
	FT	P	FT	P		
0	40	-	5	-		
1	33	91	4	66		
2	26	91	3	66		
3	18	91	2	66		
4	9	91	1	66		
5	-	91	-	66		

Legend:
FT = Family-wise Type I Error Rate (%)
MT = Marginal Type I Error Rate (%)
P = Power (%)

Results with MEM PP Threshold of 86.25%

With a posterior probability (PP) threshold of **86.25%** for the MEM approach, we observe:

# Active Baskets	Ind		Mult		Pool		MEM
	FT	P	FT	P	FT	P	
0	40	-	5	-	7	-	
1	33	91	4	66	48	48	
2	26	91	3	66	89	89	
3	18	91	2	66	99	99	
4	9	91	1	66	100	100	
5	-	91	-	66	-	100	

Legend:
FT = Family-wise Type I Error Rate (%)
MT = Marginal Type I Error Rate (%)
P = Power (%)

Results with MEM PP Threshold of 86.25%

With a posterior probability (PP) threshold of **86.25%** for the MEM approach, we observe:

# Active Baskets	Ind		Mult		Pool		MEM		
	FT	P	FT	P	FT	P	FT	MT	P
0	40	-	5	-	7	-	40	10	-
1	33	91	4	66	48	48	34	11	91
2	26	91	3	66	89	89	31	14	92
3	18	91	2	66	99	99	33	20	93
4	9	91	1	66	100	100	23	23	95
5	-	91	-	66	-	100	-	-	96

Legend:
FT = Family-wise Type I Error Rate (%)
MT = Marginal Type I Error Rate (%)
P = Power (%)

Results with MEM PP Threshold of 89.75%

With a posterior probability (PP) threshold of **89.75%** for the MEM approach, we observe:

# Active Baskets	Ind		Mult		Pool		MEM		
	FT	P	FT	P	FT	P	FT	MT	P
0	40	-	5	-	7	-	34	8	-
1	33	91	4	66	48	48	31	9	89
2	26	91	3	66	89	89	26	9	91
3	18	91	2	66	99	99	18	9	91
4	9	91	1	66	100	100	9	9	91
5	-	91	-	66	-	100	-	-	91

Legend:
FT = Family-wise Type I Error Rate (%)
MT = Marginal Type I Error Rate (%)
P = Power (%)

Results with MEM PP Threshold of 97.15%

With a posterior probability (PP) threshold of **97.15%** for the MEM approach, we observe:

# Active Baskets	Ind		Mult		Pool		MEM		
	FT	P	FT	P	FT	P	FT	MT	P
0	40	-	5	-	7	-	10	2	-
1	33	91	4	66	48	48	10	3	74
2	26	91	3	66	89	89	9	3	78
3	18	91	2	66	99	99	6	3	80
4	9	91	1	66	100	100	3	3	81
5	-	91	-	66	-	100	-	-	81

Legend:
FT = Family-wise Type I Error Rate (%)
MT = Marginal Type I Error Rate (%)
P = Power (%)

JCO-PO Paper Summary

- Methods perform very differently depending upon the given scenario
- The posterior probability threshold for MEMs can be calibrated to achieve more even performance across scenarios
- However, this does not address the larger problem of further calibrating hyperparameter values or accounting for uncertainty of the scenario we will encounter

Proposed Alternative to Selecting 1 Scenario

Instead of choosing 1 scenario, "average" over multiple scenarios and use weighted operating characteristics.

Scenarios	Basket Number				
	1	2	3	4	5
Global Null 1					
2					
3					
4					
5					
6 *Global Alternative*					

LEGEND: Null Basket Alternative Basket

Advantages of Proposed Framework

- Can use either marginal or family-wise type I error rate
- For models with hyperparameters, can optimize values and posterior probability thresholds for a given type I error rate
- Could specify different “weightings” depending on scenario(s) of interest

Hyperparameter Calibration Example

- Identify optimal calibration when equally weighting all scenarios
- Enrolling $n = 25$ per basket for 5 basket
- Assume null response of 10% and alternative response of 30% using 10,000 simulated trials for each scenario in R
- Optimize hyperparameters and posterior probability (PP) thresholds so that weighted family-wise type I error rate is $\leq 10\%$ and evaluate resulting power

Hyperparameter Calibration Results

Hyperparameter calibration is a challenging problem in the context of multi-indication studies:

Scenario	UB Value	PP Threshold
1 (5 null)	1.00	0.968
2 (4 null)	0.03	0.979
3 (3 null)	0.03	0.946
4 (2 null)	0.10	0.956
5 (1 null)	0.00	0.844
Naive Average		
Weighted		

Hyperparameter Calibration Results

Hyperparameter calibration is a challenging problem in the context of multi-indication studies:

Scenario	UB Value	PP Threshold
1 (5 null)	1.00	0.968
2 (4 null)	0.03	0.979
3 (3 null)	0.03	0.946
4 (2 null)	0.10	0.956
5 (1 null)	0.00	0.844
Naive Average	0.23	0.939
Weighted		

Hyperparameter Calibration Results

Hyperparameter calibration is a challenging problem in the context of multi-indication studies:

Scenario	UB Value	PP Threshold
1 (5 null)	1.00	0.968
2 (4 null)	0.03	0.979
3 (3 null)	0.03	0.946
4 (2 null)	0.10	0.956
5 (1 null)	0.00	0.844
Naive Average	0.23	0.939
Weighted	0.09	0.957

Family-wise Type I Error Rate Calibration Results

Comparing the extremes of choosing the ”best” scenario for information sharing (all null) to the most challenging (only 1 non-exchangeable basket) can illustrate why our more nuanced weighting approach can improve performance:

Scen. (# Null)	UB Value (PP Thresh.)	*Global Scens* T1E	Pwr	*Scenario 5* T1E	Pwr
1 (5 null)	1.00 (0.968)				
5 (1 null)	0.00 (0.844)				
Weighted	0.09 (0.957)				

Family-wise Type I Error Rate Calibration Results

Comparing the extremes of choosing the ”best” scenario for information sharing (all null) to the most challenging (only 1 non-exchangeable basket) can illustrate why our more nuanced weighting approach can improve performance:

Scen. (# Null)	UB Value (PP Thresh.)	*Global Scens* T1E	Pwr	*Scenario 5* T1E	Pwr
1 (5 null)	1.00 (0.968)	0.090	0.982		
5 (1 null)	0.00 (0.844)				
Weighted	0.09 (0.957)				

Family-wise Type I Error Rate Calibration Results

Comparing the extremes of choosing the ”best” scenario for information sharing (all null) to the most challenging (only 1 non-exchangeable basket) can illustrate why our more nuanced weighting approach can improve performance:

Scen. (# Null)	UB Value (PP Thresh.)	*Global Scens* T1E	*Global Scens* Pwr	*Scenario 5* T1E	*Scenario 5* Pwr
1 (5 null)	1.00 (0.968)	0.090	0.982	0.507	0.966
5 (1 null)	0.00 (0.844)				
Weighted	0.09 (0.957)				

Family-wise Type I Error Rate Calibration Results

Comparing the extremes of choosing the ”best” scenario for information sharing (all null) to the most challenging (only 1 non-exchangeable basket) can illustrate why our more nuanced weighting approach can improve performance:

Scen. (# Null)	UB Value (PP Thresh.)	*Global Scens* T1E	*Global Scens* Pwr	*Scenario 5* T1E	*Scenario 5* Pwr
1 (5 null)	1.00 (0.968)	0.090	0.982	0.507	0.966
5 (1 null)	0.00 (0.844)			0.094	0.911
Weighted	0.09 (0.957)				

Family-wise Type I Error Rate Calibration Results

Comparing the extremes of choosing the ”best” scenario for information sharing (all null) to the most challenging (only 1 non-exchangeable basket) can illustrate why our more nuanced weighting approach can improve performance:

Scen. (# Null)	UB Value (PP Thresh.)	*Global Scens* T1E	*Global Scens* Pwr	*Scenario 5* T1E	*Scenario 5* Pwr
1 (5 null)	1.00 (0.968)	0.090	0.982	0.507	0.966
5 (1 null)	0.00 (0.844)	0.396	0.911	0.094	0.911
Weighted	0.09 (0.957)				

Family-wise Type I Error Rate Calibration Results

Comparing the extremes of choosing the ”best” scenario for information sharing (all null) to the most challenging (only 1 non-exchangeable basket) can illustrate why our more nuanced weighting approach can improve performance:

Scen. (# Null)	UB Value (PP Thresh.)	*Global Scens* T1E	*Global Scens* Pwr	*Scenario 5* T1E	*Scenario 5* Pwr
1 (5 null)	1.00 (0.968)	0.090	0.982	0.507	0.966
5 (1 null)	0.00 (0.844)	0.396	0.911	0.094	0.911
Weighted	0.09 (0.957)	0.140	0.875	0.071	0.818

Discussion and Future Direction

- Approach is useful for designing studies and identifying optimal models for protocols
- Framework is flexible to consider any design with multiple indications, weightings of scenarios, and type I error control
- Our weighted calibration results in a better balance of type I error and power across scenarios
- Evaluating performance of MEMs in sequential monitoring with predictive probabilities for basket trials compared to a Simon 2-stage design

Sources I

Dmitrienko, A., Tamhane, A. C., and Bretz, F. (2009). *Multiple testing problems in pharmaceutical statistics*. CRC press.

Hobbs, B., Kane, M., Hong, D., and Landin, R. (2018). Statistical challenges posed by uncontrolled master protocols: sensitivity analysis of the vemurafenib study. *Annals of Oncology*.

Hobbs, B. P. and Landin, R. (2018). Bayesian basket trial design with exchangeability monitoring. *Statistics in medicine*.

Kaizer, A. M., Koopmeiners, J. S., Kane, M. J., Roychoudhury, S., Hong, D. S., and Hobbs, B. P. (2019). Basket designs: Statistical considerations for oncology trials. *JCO Precision Oncology*, 3:1–9.

Woodcock, J. and LaVange, L. M. (2017). Master protocols to study multiple therapies, multiple diseases, or both. *New England Journal of Medicine*, 377(1):62–70.

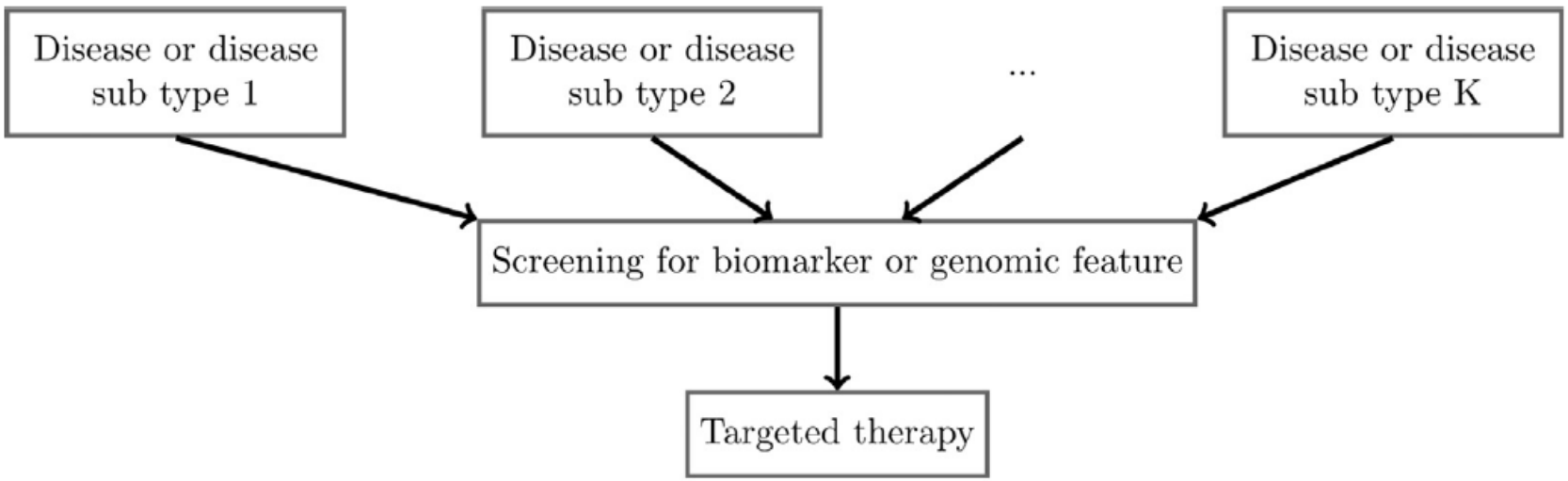
Bayesian Basket Trial Design with False Discovery Rate Control

Emily C. Zabor
Department of Quantitative Health Sciences

DIA BSWG
December 18, 2020

Cleveland Clinic

The challenge(s) of basket trials



Meyer et al. (2020). The Evolution of Master Protocol Clinical Trial Designs: A Systematic Literature Review. Clinical Therapeutics.

The challenge(s) of basket trials

https://research.bigagainstbreastcancer.org/research/translational-research

Analyzing MEMs with the basket package in R

Two fitting options:

1. `mem_mcmc`: Bayesian Metropolis-Hasting MCMC inference
2. `mem_exact`: Full Bayesian inference

Method	Return Description
`basket_pep`	Basketwise PEP matrix
`basket_map`	Basketwise MAP matrix

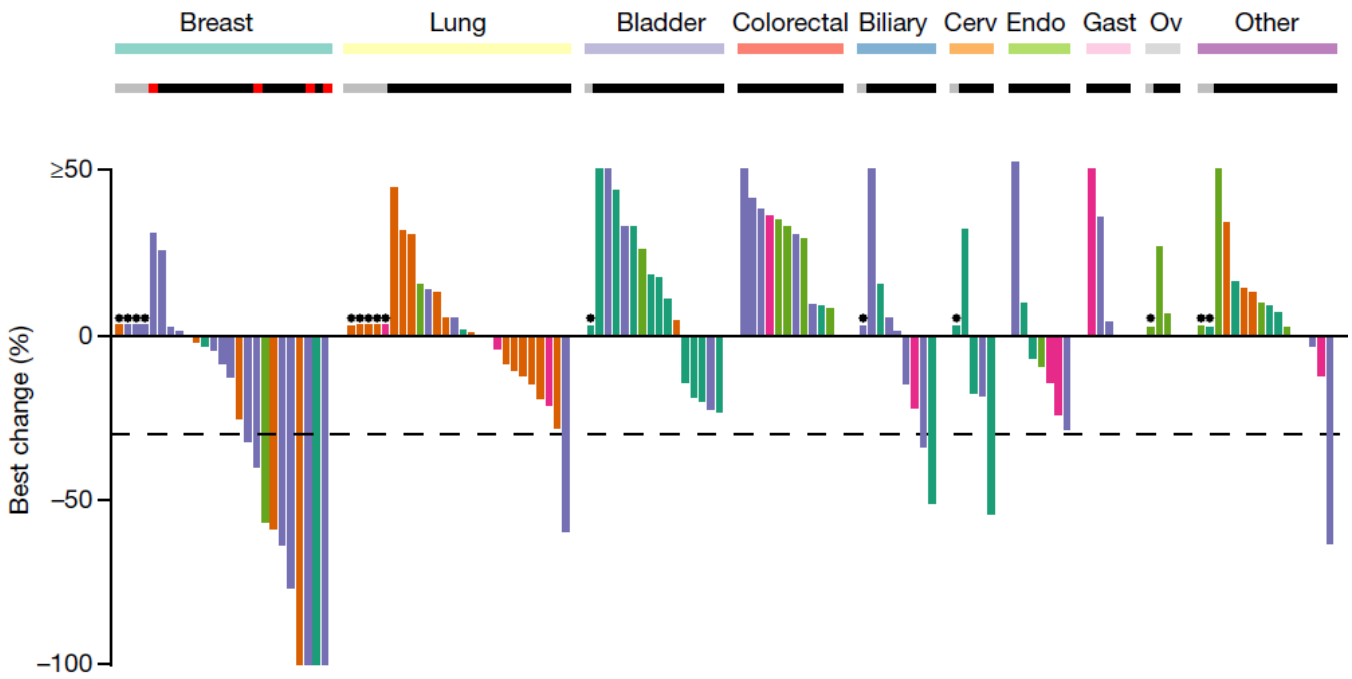
Plot Method	Return Description
`plot_pep_graph`	Network graph of the PEP matrix
`plot_pep`	Exchangeogram of the PEP matrix
`plot_map`	Exchangeogram of the MAP matrix

SUMMIT trial design

- Testing neratinib in HER2- and HER3-mutant tumors
- ORR $\leq$ 10% unacceptable, ORR $\geq$ 30% acceptable
- Independent Simon's optimal two-stage designs
  - Enroll 7 patients
  - If at least 1 response, enroll additional 11 patients
  - If 4 responses seen in 18 total patients, reject null

Hyman et al. (2018). Her kinase inhibition in patients with her2- and her3-mutant cancers. Nature 554(7691), 189-194.

SUMMIT trial results



- Breast ORR 32%

Hyman et al. (2018). Her kinase inhibition in patients with her2- and her3-mutant cancers. Nature 554(7691), 189-194.

SUMMIT data

http://www.cbioportal.org/study/summary?id=summit_2018

```
> nerat_dat
   cancer_type       enrolled responses
 1 Lung                    26         1
 2 Breast                  25         8
 3 Bladder                 18         0
 4 Colorectal              17         0
 5 Biliary tract           11         2
 6 Endometrial              8         0
 7 Cervical                 5         1
 8 Gastroesophageal         7         0
 9 Ovarian                  5         0
10 Other                   19         1
```

MCMC implementation of MEM

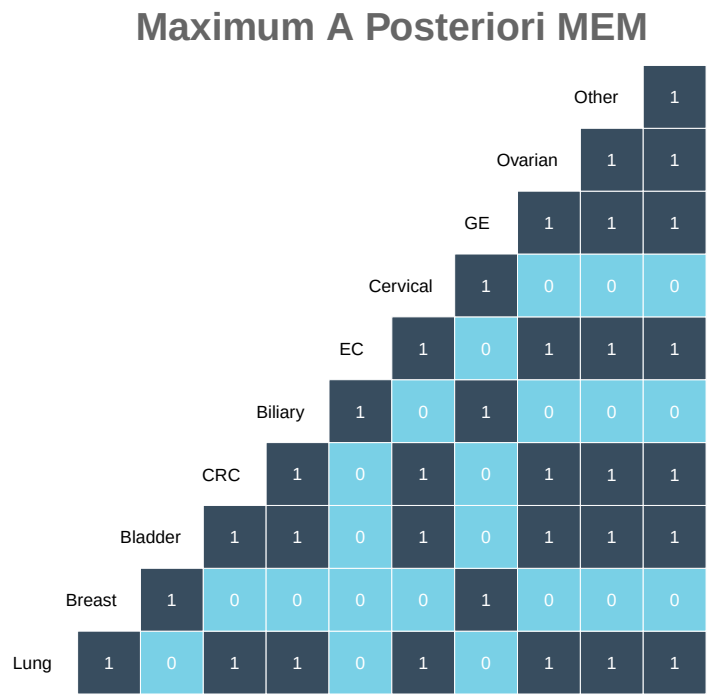
```
nerat_basket <- basket::mem_mcmc(
  responses = nerat_dat$responses,
  size = nerat_dat$enrolled,
  name = nerat_dat$cancer_type,
  p0 = 0.10,
  prior = diag(10)/(1/(1 - 0.25)) +
   matrix(0.25, nrow = 10, ncol = 10)
  )
```

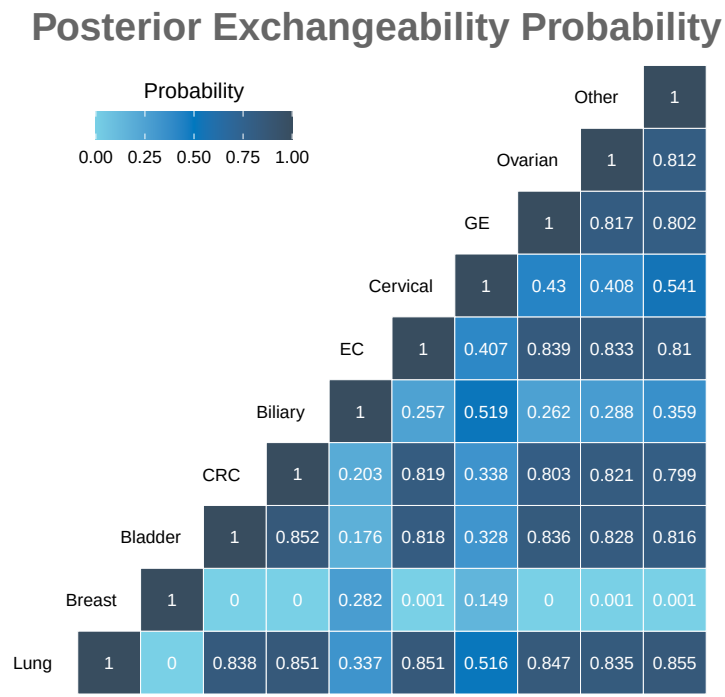
MCMC MEM implementation output

```
The Null Response Rates (alternative is greater):
                Lung Breast Bladder   CRC Biliary    EC Cervical    GE Ovarian Other
Null           0.100  0.100   0.100 0.100   0.100 0.100     0.10 0.100   0.100 0.100
Posterior Prob 0.013  0.999   0.007 0.008   0.523 0.013     0.31 0.013   0.014 0.019

Posterior Mean and Median Response Rates:
        Lung Breast Bladder   CRC Biliary    EC Cervical    GE Ovarian Other
Mean   0.037  0.312   0.031 0.031   0.139 0.034    0.095 0.034   0.034 0.039
Median 0.033  0.306   0.027 0.027   0.106 0.029    0.066 0.030   0.030 0.035

Highest Posterior Density Interval with Coverage Probability 0.95:
             Lung Breast Bladder   CRC Biliary    EC Cervical    GE Ovarian Other
Lower Bound 0.001  0.149    0.00 0.000   0.010 0.000    0.002 0.000   0.000 0.002
Upper Bound 0.079  0.484    0.07 0.071   0.335 0.077    0.285 0.077   0.078 0.084

Posterior Effective Sample Size:
   Lung Breast Bladder    CRC Biliary     EC Cervical     GE Ovarian  Other
 98.706 28.289 100.476 99.863  16.203 92.706   14.395 91.941  92.064 92.569
```

Exchangeograms of SUMMIT MAP and PEP

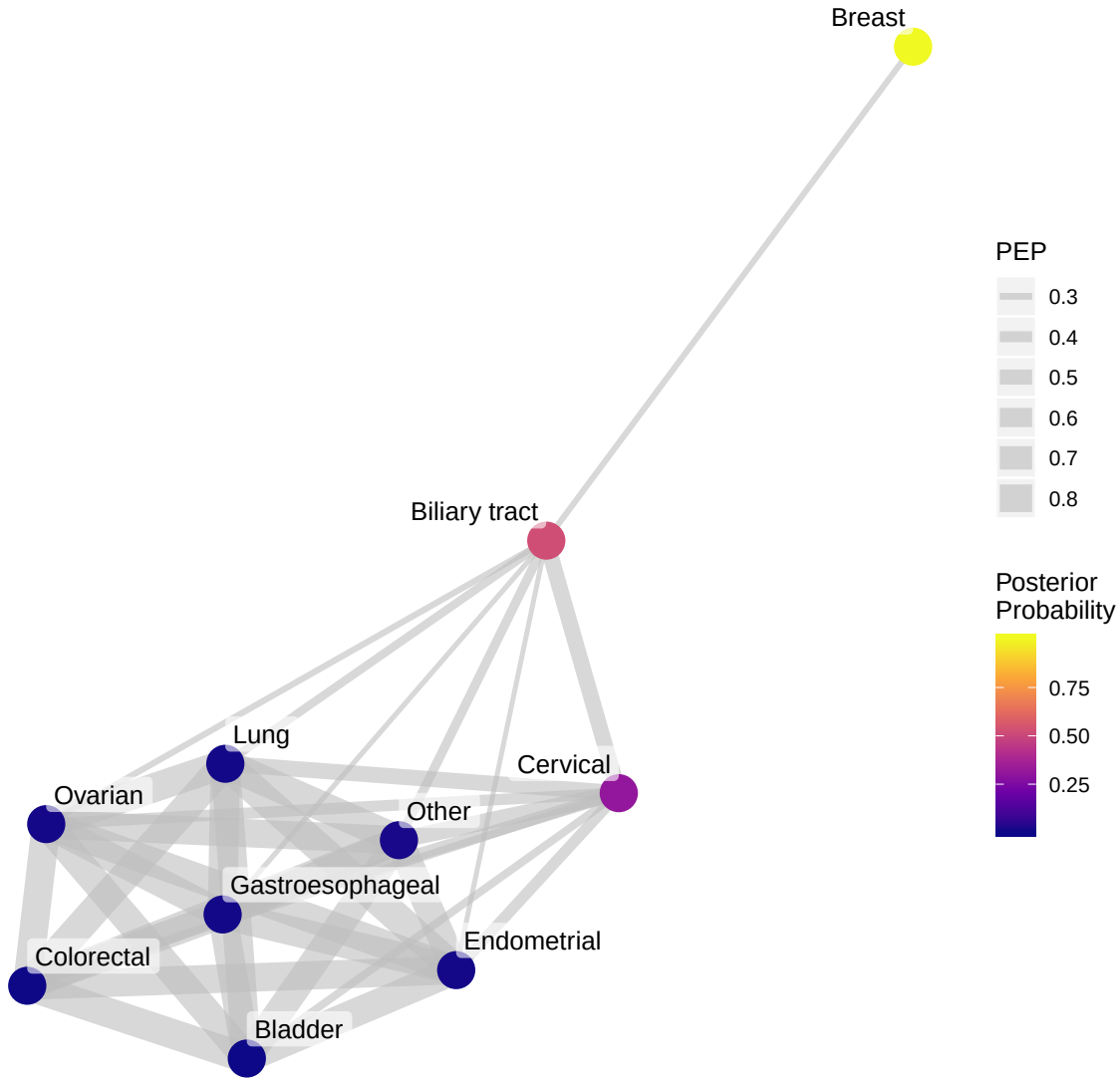
```
plot_map(nerat_basket$basket)
plot_pep(nerat_basket$basket)
```





Network graph of SUMMIT results

```
plot_pep_graph(nerat_basket, pep_cutoff = 0.25)
```



False Discovery Rate control

1. Obtain the posterior probability that each basket j exceeds the null response rate, $\Pr(\pi_j > \pi_0 | \boldsymbol{S})$
2. Order the posterior probabilities from largest to smallest, $\Pr(\pi_j > \pi_0 | \boldsymbol{S})_{(1)}, \ldots, \Pr(\pi_j > \pi_0 | \boldsymbol{S})_{(j)}$
3. For a given threshold of posterior probability, ϕ, identify the largest k such that $\Pr(\pi_j > \pi_0 | \boldsymbol{S})_{(k)} > \frac{k}{j} \times \phi$
4. Declare all baskets with posterior probability $\Pr(\pi_j > \pi_0 | \boldsymbol{S})_{(i)}$, $i = 1, \ldots, k$ to be significant at threshold ϕ

Benjamini et al. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) 57(1), 289-300.

MEM results controlling FDR at 0.05

Table 3. Basket-wise power and false positive rates (FPR) for each scenario when the overall FDR is controlled at 0.05 using the MEM approach.

	Lung	Breast	Bladder	CRC	Biliary	EC	Cervical	GE	Ovarian	Other
Power										
Global Null	–	–	–	–	–	–	–	–	–	–
Mixed Alt 1	–	0.76	–	–	–	–	–	–	–	–
Mixed Alt 2	–	0.81	–	–	0.58	–	–	–	–	–
Mixed Alt 3	–	0.82	–	–	0.61	–	0.43	–	–	–
Global Alt	0.98	0.98	0.97	0.96	0.95	0.93	0.88	0.92	0.88	0.97
FPR										
Global Null	0.059	0.054	0.049	0.042	0.052	0.039	0.027	0.047	0.028	0.045
Mixed Alt 1	0.090	–	0.091	0.100	0.094	0.083	0.071	0.088	0.082	0.102
Mixed Alt 2	0.131	–	0.128	0.126	–	0.127	0.118	0.130	0.116	0.128
Mixed Alt 3	0.146	–	0.156	0.158	–	0.154	–	0.155	0.144	0.156
Global Alt	–	–	–	–	–	–	–	–	–	–

Alt=Alternative; CRC=Colorectal; EC=Endometrial; GE=Gastroesophageal

MEM results controlling FWER at 0.05

Table 4. Basket-wise power and false positive rates (FPR) for each scenario when the overall FWER is controlled at 0.05 using the MEM approach.

	Lung	Breast	Bladder	CRC	Biliary	EC	Cervical	GE	Ovarian	Other
Power										
Global Null	–	–	–	–	–	–	–	–	–	–
Mixed Alt 1	–	0.56	–	–	–	–	–	–	–	–
Mixed Alt 2	–	0.61	–	–	0.34	–	–	–	–	–
Mixed Alt 3	–	0.63	–	–	0.38	–	0.20	–	–	–
Global Alt	0.95	0.95	0.93	0.92	0.90	0.89	0.81	0.89	0.82	0.93
FPR										
Global Null	0.010	0.008	0.011	0.008	0.008	0.004	0.008	0.003	0.006	0.005
Mixed Alt 1	0.020	–	0.018	0.024	0.020	0.014	0.014	0.019	0.019	0.025
Mixed Alt 2	0.030	–	0.031	0.034	–	0.029	0.022	0.028	0.023	0.033
Mixed Alt 3	0.046	–	0.043	0.043	–	0.038	–	0.040	0.030	0.045
Global Alt	–	–	–	–	–	–	–	–	–	–

Alt=Alternative; CRC=Colorectal; EC=Endometrial; GE=Gastroesophageal

Frequentist results controlling basket-wise error at 0.05

Table 5. Basket-wise power and false positive rates (FPR) for each scenario when the basket-wise error rate is controlled at 0.05 using the frequentist approach.

	Lung	Breast	Bladder	CRC	Biliary	EC	Cervical	GE	Ovarian	Other
Power										
Global Null	–	–	–	–	–	–	–	–	–	–
Mixed Alt 1	–	0.80	–	–	–	–	–	–	–	–
Mixed Alt 2	–	0.81	–	–	0.43	–	–	–	–	–
Mixed Alt 3	–	0.81	–	–	0.43	–	0.16	–	–	–
Global Alt	0.84	0.80	0.67	0.60	0.43	0.45	0.17	0.36	0.18	0.73
FPR										
Global Null	0.046	0.037	0.026	0.017	0.022	0.029	0.006	0.027	0.007	0.024
Mixed Alt 1	0.031	–	0.023	0.026	0.023	0.034	0.009	0.027	0.012	0.041
Mixed Alt 2	0.039	–	0.030	0.020	–	0.040	0.008	0.024	0.009	0.034
Mixed Alt 3	0.041	–	0.025	0.023	–	0.036	–	0.028	0.007	0.033
Global Alt	–	–	–	–	–	–	–	–	–	–

Alt=Alternative; CRC=Colorectal; EC=Endometrial; GE=Gastroesophageal