Analyzing Basket Trials under Multisource Exchangeability Assumptions

by Michael J. Kane, Nan Chen, Alexander M. Kaizer, Xun Jiang, H. Amy Xia, Brian P. Hobbs

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 disease settings, such as oncology wherein these trials have become popular, marginal measures of statistical evidence are difficult to interpret for sparsely enrolled subpopulations. Consequently, basket trials pose challenges to the traditional paradigm for trial design, which assumes inter-patient exchangeability. The package basket for the R programming environment facilitates the analysis of basket trials by implementing multi-source exchangeability models. By evaluating all possible pairwise exchangeability relationships, this hierarchical modeling framework facilitates Bayesian posterior shrinkage among a collection of discrete and pre-specified subpopulations. Analysis functions are provided to implement posterior inference of the response rates and all possible exchangeability relationships between subpopulations. In addition, the package can identify "poolable" subsets of and report their response characteristics. The functionality of the package is demonstrated using data from an oncology study with subpopulations defined by tumor histology.

Keywords: Bayesian analysis, basket design, hierarchical model, master protocol, oncology, patient heterogeneity

Introduction

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. Central to the design are assumptions 1) that a patient's expectation of treatment benefit can be ascertained from accurate characterization of their molecular profile and 2) that biomarker-guided treatment selection supersedes traditional clinical indicators for the studied populations, such as primary site of origin or histopathology. Thus, basket trials are designed to enroll multiple clinical subpopulations to which it is assumed that the therapy(s) in question offers beneficial efficacy in the presence of the targeted molecular profile(s). These designs have become popular as drug developers seek to conform therapeutic interventions to the individuals being treated with precision medicine and biomarker-guided therapies. Most basket trials have been conducted within exploratory settings to evaluate agent-specific estimates of tumor response. Cunanan et al. ([Cunanan et al., 2017a](#)) describe three studies implemented in oncology settings which extend the basic formulation of a basket trial to multiple targets and/or agent combinations. Most commonly uncontrolled trials, extensions have recently accommodated a wide variety of potential motivations beyond exploratory studies.

Molecularly targeted treatment strategies may not offer acceptable efficacy to all putatively promising clinical indications. Early basket trials were criticized for their reliance on basketwise analysis strategies that suffered from limited power in the presence of imbalanced enrollment as well as failed to convey to the clinical community evidentiary measures of heterogeneity among the studied clinical subpopulations, or "baskets". Acknowledging the potential for differential effectiveness among the enrolled patient subpopulations by design, heterogeneity exists as an intrinsic hypothesis in evaluations of treatment efficacy. Moreover, for rare disease settings, such as oncology wherein these trials have become popular, marginal measures of statistical evidence are difficult to interpret on the basis of individual basket-wise analyses for sparsely enrolled subpopulations. Consequently, basket trials pose specific challenges to the traditional paradigm for trial design, which assume that the patients enrolled represent a statistically exchangeable cohort.

[Hobbs and Landin](#) ([2018](#)) extended the Bayesian multisource exchangeability model (MEM) framework to basket trial design and subpopulations inference. Initially proposed by [Kaizer et al.](#) ([2017](#)), the MEM framework addressed the limitations associated with "single-source" Bayesian hierarchical models, which rely on a single parameter to determine the extent of influence, or shrinkage, from all sources. In the presence of subpopulations that arise as mixtures of exchangeable and non-exchangeable subpopulations, single-source hierarchical models (SEM) are characterized by limited borrowing, even in

the absence of heterogeneity (Kaizer et al., 2017). Moreover, when considering the effectiveness of a particular treatment strategy targeting a common disease pathway that is observed among differing histological subtypes, SEMs fail to admit statistical measures that delineate which patient subtypes should be considered "non-exchangeable" based on the observed data. By way of contrast, MEM provides a general Bayesian hierarchical modeling strategy accommodating source-specific smoothing parameters. MEMs yield multi-resolution smoothed estimators that are asymptotically consistent and accommodate both full and non-exchangeability among discrete subpopulations. The inclusion of methods for shrinkage of multiple sources is not restricted to use in basket trial master protocols, but has also been extended in the MEM framework to a sequential combinatorial platform trial design where it demonstrated improved efficiency relative to approaches without information sharing (Kaizer et al., 2018).

This paper introduces the **basket** (Chen et al., 2019) package for the R-programming environment to analyze basket trials under MEM assumptions. The main analyses conduct full posterior inference with respect to a set of response rates corresponding to the studied subpopulations. The posterior exchangeability probability (PEP) matrix is calculated, which describes the probability that any pair of baskets are exchangeable. Based on the resultant PEP, subpopulations are clustered into meta-baskets. Additionally, posterior effective sample sizes are calculated for each basket, describing the extent of posterior shrinkage achieved. Posterior summaries are reported for both "basketwise" and "clusterwise" analyses.

The package used in the examples below is available on CRAN at https://cran.r-project.org/package=basket and it fits into the general category of the "Design and Analysis of Clinical Trials" (Zhang and Zhang, 2018) focusing on uncontrolled, early-phase trial analysis. The interface is designed to be simple and will readily fit into clinical trial frameworks. It has been tested using R version 3.5 and the **basket** package version 0.9.9.

Exchangeability for Trials with Subpopulations
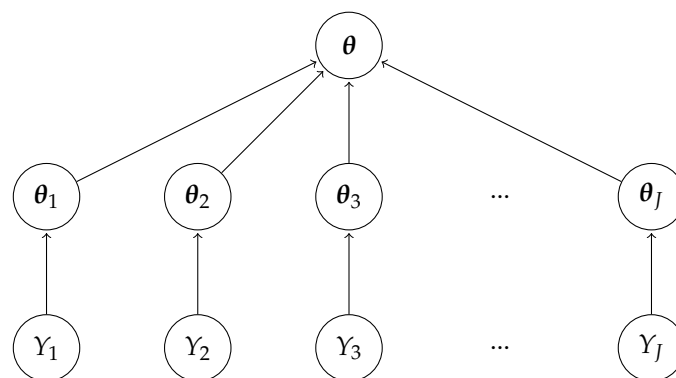
The Single-Source Exchangeability Model

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

Basket trials intrinsically include subpopulations, which require *a priori* consideration for inference. When ignored the trial simply pools patients, conducting inference with the implicit assumption of inter-patient statistical exchangeability, which can induce bias and preclude the identification of unfavorable/favorable subtypes in the presence of heterogeneity. At the other extreme, subpopulation-specific analyses assume independence. While attenuating bias, this approach suffers from low power, especially in rare subpopulations enrolling limited sample size. Bayesian hierarchical models address this polarity, facilitating information sharing by "borrowing strength" across subtypes with the intent of boosting the effective sample size of inference for individual subtypes.

Single-source exchangeability models (SEM), represent one class of Bayesian hierarchical models. In the context of a basket trial design, statistical approaches using the SEM framework rely on a single parametric distributional family to characterize heterogeneity across all subpopulations, which is computationally tractable but intrinsically reductive in characterization of heterogeneity. In the presence of both exchangeable and non-exchangeable arms, the SEM framework tends to favor the extremes of no borrowing or borrowing equally from all sources, effectively ignoring disjointed singleton subpopulations and meta-subtypes.

Consider a basket trial which enrolls patients from J subpopulations (or subtypes) ($j = 1, ..., J$),

where Y_j represents the responses observed among patients in the jth subtypes. Using i to index each patient, the SEM generally relies on model specifications that assume that patient-level responses, $Y_{i,j}$, are exchangeable Bernoulli random variables conditional on subtype-specific model parameters, e.g. θ_j. The second-level of the model hierarchy assumes that the collection of subtype-specific model parameters, $\theta_1, ..., \theta_J$, are statistically exchangeable through the specification of a common parent distribution. Figure 1 illustrates this structure, wherein each Y_j has its own subtype-specific θ_j which are further assumed exchangeable to estimate the overall θ.

Examples of SEM approaches are introduced and discussed by (Berry et al., 2010, chapter 2), Thall et al. (2003), and Berry et al. (2013), with Hobbs and Landin (2018) providing additional background on these specific SEM implementations. SEM approaches are also implemented in packages by Nia and Davison (2012) and Savage et al. (2018) and have been extended to more specialized applications in fMRI studies (Stocco, 2014), modeling clearance rates of parasites in biological organisms (Sharifi-Malvajerdi et al., 2019), modeling genomic bifurcations (Campbell and Yau, 2017), modeling ChIP-seq data through hidden Ising models (Mo, 2018), modeling genome-wide nucleosome positioning with high-throughput short-read data (Samb et al., 2015), and modeling cross-study analysis of differential gene expression (Scharpf et al., 2009).

While integrating inter-cohort information, SEMs are limited by assumptions of exchangeability among all cohorts. That is, the joint distribution $\mathbb{P}(Y_1, Y_2, ..., Y_k)$ is invariant under a permutation describing subpopulation subsets. $\mathbb{P}(Y_1, Y_2, ..., Y_k) = (Y_k, ..., Y_2, Y_1)$. SEMs are "single-source" in the sense that the model uses a single set of parameters to characterize heterogeneity such that the statistical exchangeability of model parameters is always assumed. Violations of these assumptions with analyses of response rates in clinical trials yields bias, potentially inflating the estimated evidence of an effective response rate for poorly responding cohort or minimizing the effect in effective subsets. These assumptions have resulted in poor results for frequentist power when controlling for strong type I error, leading some cancer trialists to question the utility of Bayesian hierarchical models for phase II trials enrolling discrete subtypes (Freidlin and Korn, 2013; Cunanan et al., 2017b).

The Multi-source Exchangeability Model

Limitations of SEM can be overcome through model specification devised to explicitly characterize the evidence for exchangeability among collections of subpopulations enrolling in a clinical trial. Multi-source exchangeability models (MEM) produce cohort-specific smoothing parameters that can be estimated in the presence of the data to facilitate dynamic multi-resolution smoothed estimators that reflect the extent to which subsets of subpopulations should be consider exchangeable. Shown to be asymptotically consistent, MEMs were initially proposed by Kaizer et al. (2017) for "asymmetric" cases wherein a primary data source is designated for inference in the presence of potentially non-exchangeable supplemental data sources. The framework was extended by Hobbs and Landin (2018) to the "symmetric" case wherein no single source or subtype is designated as primary (e.g., a basket trial). The symmetric MEM approach considers all possible pairwise exchangeability relationships among J subpopulations and estimates the probability that any subset of subpopulations should be considered statistically exchangeable (or poolable).

The symmetric MEM is the motivation and focus of the **basket** package. While SEMs are parameterized by a single set of parameters θ, the MEM may have up to J (the number of subtypes) sources of exchangeability with each set of data Y_j contributing to only one set of parameters. All possible combinations of exchangeability can be enumerated, denoted as K possible configurations $(\Omega_k, k = 1, ..., K)$.

(a) Model where Y_1 and Y_2 are exchangeable.   **(b)** Model where Y_1 and Y_3 are exchangeable.
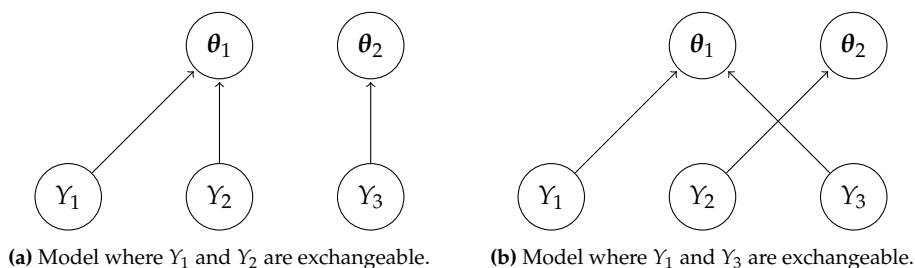
Figure 2: Two example exchangeability configurations of the MEM.

Model Description

Figure 2 depicts two possible MEMs among three subpopulations wherein at least two subpopulations are statistically exchangeable. Both examples comprise two "sources" of exchangeability for inference, with Y_1 and Y_2 combined to represent one "source" to estimate θ_1 and Y_3 to estimate θ_2 in (a) and Y_1 and Y_3 combined in (b). Implementation of **basket** considers the number of "sources" ranging from one (as in the single-source case), wherein all subtypes are pooled together, to J, the total number of subtypes. The MEM Bayesian model specification facilitates posterior inference with respect to all possible pairwise exchangeability relationships among J subpopulations. 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 set space of all possible pairwise exchangeability relationships among a collection of J discrete cohorts can be represented by a symmetric $J \times J$ matrix $\boldsymbol{\Omega}$ with element $\Omega_{ij} = \Omega_{ji} \in [0, 1]$ with value 1 (0) indicating that patients of subtype i are statistically exchangeable with (independent of) patients of subtype j. Without additional patient-level characteristics, it is assumed patients within an identical subtype are assumed to be statistically exchangeable. That is $\Omega_{ii} = 1$ for $\{i : 1, ..., J\}$. There are $K = \prod_{j=1}^{J-1} 2^j$ possible configurations of $\boldsymbol{\Omega}$, each representing one possible pairwise exchangeability relationship among the J subtypes. The framework differs fundamentally from SEM in that it allows for the existence of multiple closed subpopulations (or cliques) comprised of fully exchangeable subtypes. Therefore, following the terminology of Kaizer et al. (2017) we refer to each possible configuration of $\boldsymbol{\Omega}$ as a MEM.

For a basket trial designed to enroll a total of N patients in J baskets, let $y_{ij} = 1$ indicate the occurrence of a successful response for the ith patient enrolled in basket j, and 0 indicate treatment failure. Let n_j denote the number of patients observed in basket j and denote the total number of responses in basket j by $S_j = \sum_{i=1}^{n_j} y_{ij}$. The set $\{S_1, S_2, ..., S_J\}$ is denoted $\boldsymbol{S}$. Let $\boldsymbol{\pi} = \{\pi_1, \pi_2, ...\pi_J\}$ vectorize the set of response rates such that π_j denotes the probability of response for jth basket and $S_j \sim \text{Bin}(n_j, \pi_j)$ with prior distribution $\pi_j \sim \text{Beta}(a_j, b_j)$. Let $B()$ denote the beta function. Given an exchangeability configuration $\boldsymbol{\Omega}_j$, the marginal density of S_j follows as (see Hobbs and Landin, 2018, for details)

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

Marginal posterior inference with respect to $\pi_j \mid \boldsymbol{S}$ averages the conditional posterior of $\pi_j \mid \boldsymbol{\Omega}_j, \boldsymbol{S}$ with respect to the marginal posterior probability of $G = 2^{J-1}$ possible exchangeability configurations of $\boldsymbol{\Omega}_j$. Let $\boldsymbol{\omega} = \{\omega_1, ..., \omega_G\}$ denote the collection of vectors each of length J that collectively span the sample space of $\boldsymbol{\Omega}_j$. The marginal posterior distribution can be represented by a finite mixture density

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

where the posterior probability of exchangeability configuration ω_g given the observed data follows from Bayes' Theorem in proportion to the marginal density of the data given ω_g and its unconditional prior probability

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

Model specification for the symmetric MEM method is described in detail by Hobbs and Landin (2018).

Estimating Basketwise Exchangeability

The **basket** package computes the posterior probability that subpopulations i and j should be considered statistically exchangeable. The collection of all pairwise posterior exchangeability probabilities (PEP) is denoted in the output as the PEP matrix. Additionally, **basket** identifies the maximum *a posteriori* (MAP) multisource exchangeability model.

Let $\mathcal{O}$ denote the entire sample domain of $\boldsymbol{\Omega}$ comprised of $K = \prod_{j=1}^{J-1} 2^j$ strictly symmetric MEMs.

The PEP matrix is obtained by evaluating the union of MEMs for which $\Omega_{ij} = 1$ over the sample domain of $\mathcal{O}$,

$$\mathbb{P}(\Omega_{ij} = 1|S) = \sum_{\Omega \in \mathcal{O}} \mathbb{1}_{\{\Omega_{ij}=1\}} \mathbb{P}(\Omega|S),$$

where $\mathbb{P}(\Omega|S)$ is the product of row-wise calculations specified in Equation 3. Note that there are $K/2$https://www.overleaf.com/project/5c982c6d19014f441ddd8c2d MEM configurations in the space of $\mathcal{O}$ where $\Omega_{ij} = 1$. The MAP follows as the MEM configuration that attains maximum $Pr(\Omega \mid S)$ over $\mathcal{O}$.

Effective Sample Size

Measurement of the extent to which information has been shared across sources in the context of a Bayesian analysis is best characterized by the effective sample size (ESS) of the resultant posterior distribution Hobbs et al. (2013); Murray et al. (2015). ESS quantifies the extent of information sharing, or Bayesian "shrinkage," as the number of samples that would be required to obtain the extent of posterior precision achieved by the candidate posterior distribution when analyzed using a vague "reference" or maximum entropy prior. Calculation of the ESS in **basket** deviates from the approach suggested in Hobbs and Landin (2018), which is sensitive to heavy-tailed posteriors. Robustness is introduced with **basket** through beta distributional approximation, which yields more conservative estimates of ESS. Specifically, the simulated annealing algorithm (implemented with GenSA package Yang Xiang et al. (2013)) is used to identify the parametric beta distribution with minimal Euclidean distance between the interval boundaries obtained from the posterior estimated HPD interval and the corresponding beta $1-$hpd_alpha Bayesian credible interval. Shape parameters attained from the "nearest" parametric beta distribution are summed to yield estimates of posterior ESS for each basket and cluster.

Posterior Probability

Basket trials are devised for the purpose of testing the hypothesis that a targeted treatment strategy achieves sufficiently promising activity among a partition of the targeted patient population. The MEM framework acknowledges the potential for heterogeneity with respect to the effectiveness of the enrolled patient subpopulations or baskets. 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 a null value π_0 is computed by the weighted average of cdfs for all possible exchangeability configurations. **basket** implements this computation and allows for subpopulation-specific values of the null hypothesis, π_0, which quantify differing benchmarks for effectiveness among the studied baskets. Note that this feature accommodates basket formulation on the basis of varying levels of clinical prognosis.

Package Overview

The **basket** package facilitates implementation of the binary, symmetric multi-source exchangeability model with posterior inference arising with both exact computation and Markov chain Monte Carlo sampling. The user is required to input vectors that describe the number of samples (size) and observed successes (responses) corresponding to each subpopulation (or basket). Analysis output includes full posterior samples, highest posterior density (HPD) interval boundaries, effective sample sizes (ESS), mean and median posterior estimates, posterior exchangeability probability matrices, and the maximum *a posteriori* MEM. Subgroups can be combined into meta-baskets, or clusters, by setting logical argument cluster_analysis to TRUE. Cluster analyses use graphical clustering algorithms implemented with the igraph package.

A specific clustering algorithm needs to be specified via argument cluster_function. The cluster_function is a user defined function that first creates a graph using the MAP, then assigns the baskets to discrete clusters using one of the community detection algorithms implemented in the igraph package. The default value of cluster_function is cluster_membership, a function defined in the **basket** package that implements cluster analysis based on the "cluster_louvain" method. Users can define their own cluster_function using different clustering methods in the **igraph** package. cluster_analysis is set to FALSE by default. The package includes similar calculations, summaries, and visualization for "clusterwise" and "basketwise" results. Additionally, plotting tools are provided to visualize basket and cluster densities as well as their pairwise exchangeability.

Analysis requires the specification of beta shape parameters (shape1 and shape2) for the prior distributions of the basketwise response probabilities π_j. Shape parameter arguments may be specified

as single positive real values, by which identical prior distributions are assumed for all π_j, or as vectors of length J with each pair of shape1 and shape2 values corresponding to each basket. Arguments shape1 and shape2 assume values 0.5 by default characterizing prior distributions with the effective sample size of 1 patient for each π_j.

The user must additionally specify the symmetric matrix of prior exchangeability probabilities (prior). The model assumes that exchangeable information is contributed among patients enrolling into a common basket. Thus, all diagonal entries of prior must assume value 1. Off-diagonal entries, however, quantify the *a priori* belief that each pair of subpopulations represents an exchangeable unit. Thus, off-diagonal cells of prior may assume any values on the unit interval. The **basket** package assumes the "reference" prior proposed by Hobbs and Landin (2018) as the default setting for which all off-diagonal cells assume prior probability 0.5, and thus are unbiased with respect to exchangeability in the absence of the data.

Evidence for sufficient activity is reported by basket and cluster as posterior probabilities. Posterior probability calculations require the further specification of either a null response rate or vector of null response rates corresponding to each basket (p0 set to 0.15 by default) as well as the direction of evaluation (alternative set to "greater" by default). Additionally, summary functions report the posterior estimates by basket and cluster. The highest posterior density (HPD) is calculated for a given a level of probabilistic significance (hpd_alpha set to 0.05 by default).

Bayesian computation is implemented by two methods: the exact method (mem_exact() function) and the Markov chain Monte Carlo (MCMC) sampling method (mem_mcmc() function). mem_mcmc() is the preferred method. mem_exact() provides slightly more precise estimates than the former but scales poorly in number of baskets. The discrepancy in precision between exact and sampling-based implementations is easily controlled by specifying a larger number of MCMC iterations (num_iter set to 2e+05 by default) in mem_mcmc().

The Exact Method and the MCMC Method

Implementation of mem_exact() conducts posterior inference through enumeration of the entire sample domain of MEMs, denoted $\mathcal{O}$ above. Facilitating precise calculation of the posterior estimators, mem_exact() is computationally feasible only in the presence of a small number of subpopulations. Increasing the size of J increases the number of configurations in $\mathcal{O}$ by order of $\mathcal{O}(2^{J^2})$. Thus, the exact computation is impractical for large values of J. We recommend its use for $J < 7$.

Our MCMC sampling method, formulated from the Metropolis algorithm (see e.g. Gelman et al., 2013), extends the model's implementation to larger collections of subpopulations, which currently accommodates more than $J = 20$ baskets. Specifically, MCMC sampling is used to approximate the posterior distribution $\mathbb{P}(\Omega_j = \omega_g | S)$. Implementation of mem_mcmc() requires the specification of an initial MEM matrix (initial_mem) used as the starting point for Ω from which to initiate the Metropolis algorithm. Argument initial_mem is set to round(prior -0.001) by default, which for the default setting of prior yields the identity matrix.

The MCMC algorithm proceeds in iterative fashion with each step selecting a random number of cells of Ω to flip from 0 to 1 or from 1 to 0 to produce a new candidate MEM which we denote Ω^*. Acceptance criteria for the candidate Ω^* compares the marginal posterior density of Ω^* and its unconditional prior distribution with respect to the last accepted MEM matrix configuration. Denote the sum of log marginal posterior density and prior distribution with new candidate MEM configuration by D^* and previously accepted configuration by D_0, respectively. If $D^* - D_0 \geq 0$, the candidate configuration is accepted. Otherwise, the new configuration is accepted randomly with probability $\exp(D^* - D_0)$. For each sampled Ω configuration, π_j, is sample from its conditional posterior distribution for all $j = 1, ..., J$.

The algorithm initiates with a burn-in period (mcmc_burnin set to 50,000 by default). Discarding the burn-in samples, PEP calculation with mem_mcmc() evaluates the distribution of sampled MEMs, reporting for all basket pair combinations the proportion of samples that identify basket i as exchangeable with basket j. The MAP calculation reports the posterior mode or most frequently sampled MEM. Bayesian computation facilitated by mem_mcmc() scales MEM analyses to more than 20 baskets. Specification of the size of the MCMC iterations (num_iter) is pivotal to attaining precise estimates of the resultant posterior quantities. Our investigations support the default value of 2e+05 as a practical lower bound. In practice, one may gradually increase the number of the MCMC iterations until the resultant PEP matrix converges to stable values.

Method	Return Description
basket_pep	Basketwise PEP matrix
basket_map	Basketwise maximum *a posteriori* probability (MAP) matrix
cluster_baskets	Basket assignments for each cluster
cluster_pep	Clusterwise PEP matrix
cluster_map	Clusterwise MAP matrix

Table 1: MEM model accessor functions.

MEM Data Structure and Associated Methods

Analysis functions `mem_mcmc()` and `mem_exact()` are parameterized almost identically, with the former requiring extra arguments that control the MCMC algorithm: the current seed (for reproducibility), the length of burn-in and number of MCMC iterations for computation of posterior quantities, and an initial MEM matrix from which to start the algorithm. Function arguments are specified with reasonable default values for implementation of either analysis type. Both functions return a common list data structure. Both are derived from an abstract S3 "exchangeability_model" class with concrete type "mem_mcmc" or "mem_exact" depending on which function generated the analysis. The two data structures differ only by extra elements included with "mem_mcmc" objects to control implementation of the MCMC algorithm. For convenience, and to promote using "mem_mcmc" by default, a wrapper function `basket()` was created. The `method` argument allows the user to specify the analysis function as either MCMC (via "mcmc") or exact (via "exact"). By default the argument is set to "mcmc".

MEM or "exchangeability" objects are composed of named elements. The first, `"call"` is the expression used to generate the analysis. Second is the `"basket"` element, which is a list with concrete class `mem_basket`, derived from the `mem` abstract class. Basket reports posterior estimates of trial subpopulations including the PEP, HPD interval, posterior probability, ESS, and other distribution characteristics. The `"cluster"` element comprises a list with concrete class `mem_cluster` and abstract class `mem` which contains posterior estimates for clusters rather than baskets. In addition to these three elements, an `mem_mcmc` object will also contain the seed used to generate the results. This value can be used to reproduce subsequent analyses.

Because they are relatively complex, a `summary` function is implemented to summarize the components relevant to exchangeability models for trial analysis. The `summary.exchangeability_model()` method returns an object of type `"mem_summary"`. A `print.mem_summary()` method is provided for a user-readable summary of the trial. Because there is little distinction between an `exchangeability_model` object and its summary, `print.exchageability_model()` method prints the summary object.

The `mem_summary` object provides access to the overall study characteristic. Accessor methods are also provided to extract other key information from the analysis objects at both the basket and cluster levels. These functions and their descriptions are given in Table 1. In addition, a complete MEM analysis is computationally intensive; altering the null response rate need not imply rerunning the entire analysis. To facilitate partial analysis updates under a new null (argument `p0`), the `update_p0()` function is provided. Likewise samples can be drawn from the posterior distribution of the basket and cluster models using the `sample_posterior()` function.

Visualizations

Two types of functions are provided for visualizing the results of an MEM analysis, both of which are supported at basket and cluster levels of inference. Density plotting is available with the `plot_density()` functions, which produce graphs depicting the posterior distributions of response probabilities at the basket and cluster level. Additionally, functions for visualizing exchangeability relationships are provided in a manner similar to correlograms. Since the values visualized are exchangeability, rather than correlation, we have termed these plots *exchangeograms*. These can be plotted for PEP and MAP matrices using the `plot_pep()` and `plot_map()` functions, respectively. A network graphical visualization integrating the resultant PEP and posterior probability is provided via function `plot_PEP_graph()`.

Case Study: The Vemurafenib Basket Trial

The "Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations" study (Hyman et al., 2015), enrolled patients into predetermined baskets that were determined by organ site with

Basket	Enrolled	Evaluable	Responses	Response Rate
NSCLC	20	19	8	0.421
CRC (vemu)	10	10	0	0.000
CRC (vemu+cetu)	27	26	1	0.038
Bile Duct	8	8	1	0.125
ECD or LCH	18	14	6	0.429
ATC	7	7	2	0.286

Table 2: Vemurafenib trial enrollment and responses.

primary end point defined by Response Evaluation Criteria in Solid Tumors (RECIST), version 1.1 (Eisenhauer et al., 2009) or the criteria of the International Myeloma Working Group (IMWG) (Durie et al., 2006). Statistical evidence for preliminary clinical efficacy was obtained through estimation of the organ-specific objective response rates at 8 weeks following the initiation of treatment. This section demonstrates the implementation of **basket** through analysis of six organs comprising non–small-cell lung cancer (NSCLC), cholangiocarcinoma (Bile Duct), Erdheim–Chester disease or Langerhans'-cell histiocytosis (ECD or LCH), anaplastic thyroid cancer (ATC), and colorectal cancer (CRC) which formed two cohorts. Patients with CRC were initially administered vemurafenib. The study was later amended to evaluate vemurafenib in combination with cetuximab for CRC which comprised a new basket. Observed outcomes are summarized in Table 2 by basket. Included in the **basket** package, the dataset is accessible in short `vemu_wide` as well as long formats `vemu`.

Inspection of Table 2 reveals heterogeneity among the studied baskets. CRC (vemu), CRC (vemu+cetu), and Bile Duct had relatively low response rates when compared to other baskets, suggesting that patients presenting the BRAF V600 mutation may not yield exchangeable information for statistical characterization of the effectiveness of the targeted therapy. Therefore, the MEM framework is implemented to measure the extent of basketwise heterogeneity and evaluate the effectiveness of the targeted therapy on the basis of its resultant multi-resolution smoothed posterior distributions. Hobbs et al. (2018) present a permutation study which extends the evaluation of heterogeneity to evaluate summaries of patient attributes reported in Table 1 of the aforementioned trial report. This case study reports posterior probabilities evaluating the evidence that the response probability for each organ-site exceeds the null rate of $p0 = 0.25$.

The analysis can be reproduced by loading the `vemu_wide` data, which is included with the package. The data set includes the number of evaluable patients (column `evaluable`), the number of responding patients (column `responders`), and the associated baskets for the respective results (column `baskets`). The model is fit by passing these values to the `basket()` function along with an argument specifying the null response rate of 0.25 for evaluation of each basket. The results are shown by passing the fitted model object to the `summary()` function. Code to perform the analysis as well as produce the output is shown below.

```
library(basket)
    data(vemu_wide)
    vm <- basket(vemu_wide$responders, vemu_wide$evaluable,
    vemu_wide$baskets, p0 = 0.25, cluster_analysis = TRUE)
    summary(vm)


  -- The MEM Model Call --------------------------------------------------


mem_mcmc(responses = responses, size = size, name = name, p0 = p0,
    shape1 = shape1, shape2 = shape2, prior = prior, hpd_alpha = hpd_alpha,
    alternative = alternative, mcmc_iter = mcmc_iter, mcmc_burnin = mcmc_burnin,
    initial_mem = initial_mem, seed = seed, cluster_analysis = cluster_analysis,
    call = call, cluster_function = cluster_function)


  -- The Basket Summary --------------------------------------------------


The Null Response Rates (alternative is greater):
              NSCLC CRC (vemu) CRC (vemu+cetu) Bile Duct ECD or LCH    ATC
Null          0.250     0.250            0.25      0.250       0.25  0.250
Posterior Prob 0.972    0.003            0.00      0.225       0.97  0.891


Posterior Mean and Median Response Rates:
      NSCLC CRC (vemu) CRC (vemu+cetu) Bile Duct ECD or LCH    ATC
```

```
Mean    0.394       0.055           0.053    0.148      0.394 0.358
Median 0.392        0.046           0.045    0.097      0.391 0.361


Highest Posterior Density Interval with Coverage Probability 0.95:
            NSCLC CRC (vemu) CRC (vemu+cetu) Bile Duct ECD or LCH  ATC
Lower Bound 0.242       0.00            0.001     0.005      0.238 0.17
Upper Bound 0.550       0.13            0.122     0.403      0.551 0.56


Posterior Effective Sample Size:
  NSCLC CRC (vemu) CRC (vemu+cetu) Bile Duct ECD or LCH    ATC
 37.254     49.039          54.514    10.528      36.148 21.786


-- The Cluster Summary --------------------------------------------------------


Cluster 1
 "CRC (vemu)" "CRC (vemu+cetu)" "Bile Duct"
Cluster 2
 "NSCLC" "ECD or LCH" "ATC"


The Null Response Rates (alternative is greater):
               Cluster 1 Cluster 2
Null              0.250     0.250
Posterior Prob    0.076     0.944


Posterior Mean and Median Response Rates:
        Cluster 1 Cluster 2
Mean       0.085     0.382
Median     0.057     0.382


Highest Posterior Density Interval with Coverage Probability 0.95:
            Cluster 1 Cluster 2
Lower Bound    0.000     0.221
Upper Bound    0.313     0.559


Posterior Effective Sample Size:
 Cluster 1 Cluster 2
     9.786     30.12
```

Bayesian MEM analysis using the MCMC sampler with reference prior distribution for exchangeability identifies the most likely MEM to be comprised of two closed subgraphs (or meta-baskets). Cluster 1 consists of CRC (vemu) with CRC (vemu+cetu) and BD, while cluster 2 is comprised of NSCLC, ECD or LCH, and ATC. Cluster 1 results in an estimated posterior mean response rate of 0.087. The posterior probability that baskets assigned to cluster 1 exceed the null response rate of 0.25 is only 0.082. Conversely, attaining a posterior probability of 0.944 and posterior mean of 0.382, indications identified in cluster 2 demonstrate more promising indications of activity. Figures 3a and 3b depict full posterior distributions of response probabilities for each basket and cluster produced by the plot_density() function.

```
        plot_density(vm, type = "basket")
        plot_density(vm, type = "cluster")
```

The resultant posterior probability of each pairwise exchangeability relationship (PEP) is summarized with the basket_pep() function and depicted in Figure 4 by application of the plot_pep() function. The results demonstrate that the posterior exchangeability between the high-response baskets is higher than that of the lower responding baskets. For example, the posterior probability that NSCLC and ED.LH patients are exchangeable with respect to evaluating their response to Vemurafenib is 0.938. Similarly, the analysis resulted in PEPs of 0.86 for the pairwise relationships between NSCLC with ATC and ED.LH with ATC, suggesting that these indications can be averaged. The study provided strong support to conclude that vemurafenib is identically ineffective among CRC (vemu) and CRC (vemu+cetu) subtypes with PEP = 0.92. The effectiveness of BD was identified as marginally exchangeable with CRC (vemu) and CRC (vemu+cetu) with PEP = 0.64 and 0.63, respectively. Conversely, both NSCLC and ED.LH resulted in PEPs of 0 for each CRC basket, demonstrating strong evidence of differential activity among these indications. Thus, definitive trials devised to estimate population-averaged effects should not expect these subtypes to comprise statistically exchangeable
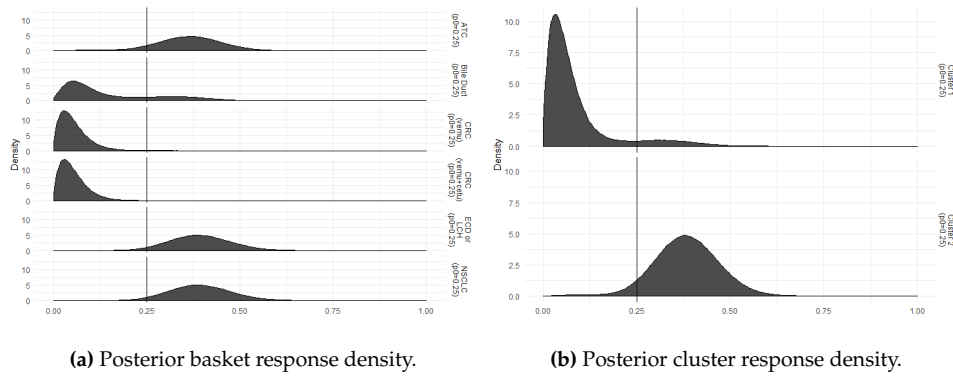
(a) Posterior basket response density.

(b) Posterior cluster response density.

Figure 3: Posterior distributions of the MEM analysis.

patients.

Figure 5 provides a network graphical representation of the results from analysis of the Vemurafenib study. This graph is generated using the `plot_pep_graph()` function. Nodes represent individual baskets. A node's color depicts the Bayesian evaluation of the null hypothesis that the posterior probability that the objective response rate exceeds 0.25 for the corresponding basket. Edge thickness between any pair of baskets is determined by PEP. Edges with shorter length and thicker width denote basket pairs with higher magnitudes of pairwise posterior exchangeability. For example, baskets ATC, NSCLC, and ECD or LCH, depicted with yellow colored nodes, resulted in higher poster probability when compared to the other three baskets. The relatively thick edges between these baskets confer their large PEP values, suggesting that these indications can be averaged.

```
basket_pep(vm)
plot_pep(vm$basket)
plot_pep_graph(vm)
```

```
                NSCLC CRC (vemu) CRC (vemu+cetu) Bile Duct ECD or LCH   ATC
NSCLC           1.000      0.002           0.000     0.231      0.938 0.866
CRC (vemu)      0.002      1.000           0.917     0.643      0.002 0.068
CRC (vemu+cetu) 0.000      0.917           1.000     0.626      0.000 0.031
Bile Duct       0.231      0.643           0.626     1.000      0.243 0.536
ECD or LCH      0.938      0.002           0.000     0.243      1.000 0.861
ATC             0.866      0.068           0.031     0.536      0.861 1.000
```
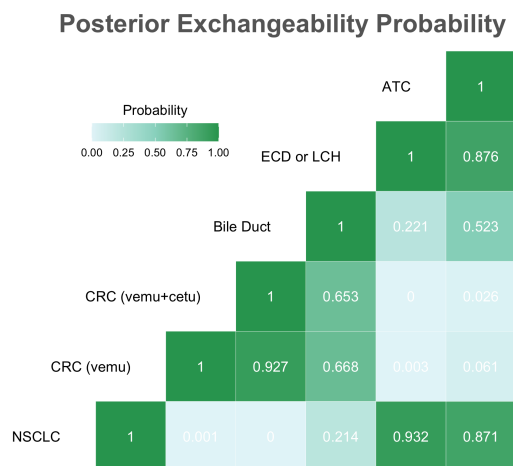

Figure 4: The exchangeogram depicting PEP resulting from analysis of the Vemurafenib study.
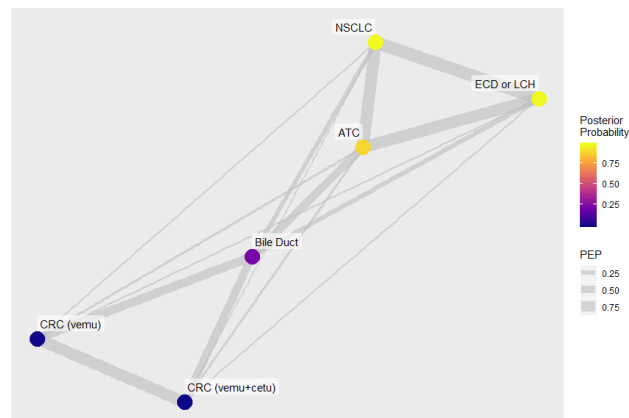
Figure 5: Network graphical representation of PEP resulting from analysis of the Vemurafenib study.

Summary

With the emergence of molecularly targeted therapies, contemporary trials are devised to enroll potentially heterogeneous patient populations defined by a common treatment target. Consequently, characterization of subpopulation heterogeneity has become central to the design and analysis of clinical trials, in oncology in particular. By partitioning the study population into subpopulations that comprise potentially non-exchangeable patient cohorts, the basket design framework can be used to study treatment heterogeneity in a prospective manner. When applied in this context, the Bayesian multisource exchangeability model (MEM) methodology refines the estimation of treatment effectiveness to specific subpopulations. Additionally, the MEM inferential strategy objectively identifies which patient subpopulations should be considered exchangeable and to what extent.

This article introduced the R package **basket** as well as demonstrated its implementation for basket trial analysis using the MEM methodology. An oncology case study using data acquired from a basket trial was presented and used to demonstrate the main functionality of the package. The **basket** package is the first available software package implementing Bayesian analysis with the MEM. The package is being actively maintained and used in ongoing trials.

Bibliography

S. M. Berry, B. P. Carlin, J. J. Lee, and P. Müller. *Bayesian Adaptive Methods for Clinical Trials*. Chapman and Hall/CRC Press, Boca Raton, FL, 2010. [p361]

S. M. Berry, K. R. Broglio, S. Groshen, and D. A. Berry. Bayesian hierarchical modeling of patient subpopulations: Efficient designs of phase II oncology clinical trials. *Clinical Trials*, 10(5):720–734, 2013. [p361]

K. R. Campbell and C. Yau. Probabilistic modeling of bifurcations in single-cell gene expression data using a bayesian mixture of factor analyzers. *Wellcome open research*, 2, 2017. doi: 10.12688/wellcomeopenres.11087.1. URL http://dx.doi.org/10.12688/wellcomeopenres.11087.1. [p361]

N. Chen, B. Hobbs, A. Kaizer, and M. J. Kane. *basket: Basket Trial Analysis*, 2019. R package version 0.9.2. [p360]

K. M. Cunanan, M. Gonen, R. Shen, D. M. Hyman, G. J. Riely, C. B. Begg, and A. Iasonos. Basket trials in oncology: A trade-off between complexity and efficiency. *Journal of Clinical Oncology*, 35(3):271–273, 2017a. doi: 10.1200/JCO.2016.69.9751. URL https://doi.org/10.1200/JCO.2016.69.9751. PMID: 27893325. [p359]

K. M. Cunanan, A. Iasonos, R. Shen, D. M. Hyman, G. J. Riely, M. Gönen, and C. B. Begg. Specifying the true-and false-positive rates in basket trials. *JCO Precision Oncology*, 1:1–5, 2017b. [p361]

B. G. Durie, J. Harousseau, J. Miguel, J. Blade, B. Barlogie, K. Anderson, M. Gertz, M. Dimopoulos, J. Westin, P. Sonneveld, et al. International uniform response criteria for multiple myeloma. *Leukemia*, 20(9):1467, 2006. [p366]

E. A. Eisenhauer, P. Therasse, J. Bogaerts, L. H. Schwartz, D. Sargent, R. Ford, J. Dancey, S. Arbuck, S. Gwyther, M. Mooney, et al. New response evaluation criteria in solid tumours: revised recist guideline (version 1.1). *European journal of cancer*, 45(2):228–247, 2009. [p366]

B. Freidlin and E. Korn. Borrowing information across subgroups in phase II trials: Is it useful? *Clinical Cancer Research*, 19:1326–1334, 2013. [p361]

A. Gelman, J. B. Carlin, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis*. Chapman and Hall/CRC Press, Boca Raton, FL, 3rd edition, 2013. [p364]

B. Hobbs, M. Kane, D. Hong, and R. Landin. Statistical challenges posed by uncontrolled master protocols: sensitivity analysis of the vemurafenib study. *Annals of Oncology*, 29(12):2296–2301, 2018. [p366]

B. P. Hobbs and R. Landin. Bayesian basket trial design with exchangeability monitoring. *Statistics in medicine*, 37(25):3557–3572, 2018. [p359, 361, 362, 363, 364]

B. P. Hobbs, B. P. Carlin, and D. J. Sargent. Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials*, 10:430–440, 2013. [p363]

D. M. Hyman, I. Puzanov, V. Subbiah, J. E. Faris, I. Chau, J.-Y. Blay, J. Wolf, N. S. Raje, E. L. Diamond, A. Hollebecque, et al. Vemurafenib in multiple nonmelanoma cancers with braf v600 mutations. *New England Journal of Medicine*, 373(8):726–736, 2015. [p365]

A. M. Kaizer, J. S. Koopmeiners, and B. P. Hobbs. Bayesian hierarchical modeling based on multisource exchangeability. *Biostatistics*, 19(2):169–184, 2017. [p359, 360, 361, 362]

A. M. Kaizer, B. P. Hobbs, and J. S. Koopmeiners. A multi-source adaptive platform design for testing sequential combinatorial therapeutic strategies. *Biometrics*, 74(3):1082–1094, 2018. [p360]

Q. Mo. *iSeq: Bayesian Hierarchical Modeling of ChIP-seq Data Through Hidden Ising Models*, 2018. R package version 1.34.0. [p361]

T. A. Murray, B. P. Hobbs, and B. P. Carlin. Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. *Annals of Applied Statistics*, 9 (3):1549–1570, 2015. [p363]

V. P. Nia and A. C. Davison. High-dimensional bayesian clustering with variable selection: The R package bclust. *Journal of Statistical Software*, 47(5):1–22, 2012. URL http://www.jstatsoft.org/v47/i05/. [p361]

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019. URL https://www.R-project.org/. [p]

R. Samb, K. Khadraoui, P. Belleau, A. Deschênes, L. Lakhal-Chaieb, and A. Droit. Using informative multinomial-dirichlet prior in a t-mixture with reversible jump estimation of nucleosome positions for genome-wide profiling. *Statistical Applications in Genetics and Molecular Biology*, 14, 2015. doi: 10.1515/sagmb-2014-0098. URL https://doi.org/10.1515/sagmb-2014-0098. [p361]

R. Savage, E. Cooke, R. Darkins, and Y. Xu. *BHC: Bayesian Hierarchical Clustering*, 2018. R package version 1.34.0. [p361]

R. B. Scharpf, H. Tjelmeland, G. Parmigiani, and A. Nobel. A bayesian model for cross-study differential gene expression. *JASA*, 2009. URL 10.1198/jasa.2009.ap07611. [p361]

S. Sharifi-Malvajerdi, F. Zhu, C. B. Fogarty, M. P. Fay, R. M. Fairhurst, J. A. Flegg, K. Stepniewska, and D. S. Small. Malaria parasite clearance rate regression: an r software package for a bayesian hierarchical regression model. *Malaria Journal*, 18(1):4, Jan 2019. ISSN 1475-2875. doi: 10.1186/s12936-018-2631-8. URL https://doi.org/10.1186/s12936-018-2631-8. [p361]

A. Stocco. Coordinate-based meta-analysis of fmri studies with r. *R Journal*, 6(2), 2014. [p361]

P. Thall, J. Wathen, B. Bekele, R. Champlin, L. Baker, and R. Benjamin. Hierarchical bayesian approaches to phase II trials in diseases with multiple subtypes. *Statistics in Medicine*, 22:763–780, 2003. [p361]

Yang Xiang, S. Gubian, B. Suomela, and J. Hoeng. Generalized simulated annealing for efficient global optimization: the GenSA package for R. *The R Journal Volume 5/1, June 2013*, 2013. URL https://journal.r-project.org/archive/2013/RJ-2013-002/index.html. [p363]

E. Zhang and H. G. Zhang. CRAN Task View: Clinical Trial Design, Monitoring, and Analysis. https://CRAN.R-project.org/view=ClinicalTrials, 2018. Version 2018-06-18. [p360]

Michael J. Kane
School of Public Health
Biostatistics Department
Yale University
New Haven, CT, USA
E-mail: michael.kane@yale.edu

Nan Chen
MD Anderson Cancer Center
The University of Texas
Houston, TX, USA
E-mail: nchen2@mdanderson.org

Alex Kaizer
Colorado School of Public Health
Department of Biostatistics and Informatics
University of Colorado-Anschutz Medical Campus
Aurora, CO, USA
E-mail: alex.kaizer@cuanschutz.edu

Xun Jiang
Amgen Inc. Thousand Oaks, CA, USA
E-mail: xunj@amgen.com

H. Amy Xia
Biostatistics and Design & Innovation
Amgen Inc. Thousand Oaks, CA, USA
E-mail: hxia@amgen.com

Brian Hobbs
Taussig Cancer Center
The Cleveland Clinic
Cleveland, OH, USA
E-mail: HobbsB@ccf.org