Bayesian Mixture Response Analysis for Pre-clinical In Vivo Anti-Cancer Drug Efficacy Investigations

David L Gold[1] and Ningning Chen[2]
[1]Bristol Myers-Squibb, Princeton NJ.
[2]Apple Inc., Cupertino CA

Abstract

Recent and notable immunotherapeutic successes for the treatments of melanoma, lung, and renal cancers have catalyzed wide-spread development of novel immunotherapies and combinations for a variety of cancers. Pre-clinical identification of efficacious and synergistic drug combinations *in vivo* remains perplexing. Among the many experimental hurdles, inbred syngeneic replicate mice have been observed to display categorically heterogeneous tumor responses, whereby subgroups of inbred mice can show different outcomes: resistance, partial response, or complete response, to the same immuno-interventions. Statistical analysis is therefore complicated by the various ways that a drug combination can be superior to single agents.

We applied Bayesian statistics to evaluate drug combinations in syngeneic mouse tumor experiments, in the presence of heterogeneous categorical responses to treatment. Our Bayesian procedure can be applied to compare drug combinations to monotherapies in multiple endpoints simultaneously, thereby inferring multiple possible ways that a drug combination can be superior to monotherapies. We propose our Bayesian statistical analysis for evaluating immunotherapeutic drug combinations in syngeneic mice.

Key Words: Bayesian, Tumor, Mouse, Drug Combination, Immunotherapy, Synergy, Categorical, Mixture

1. Introduction

Immuno-oncology (IO) drugs are designed to modify host immune responses against cancerous cells. Notable IO successes have been realized for the clinical treatments of multiple cancer types, encouraging further advances (1). As a relatively new paradigm for anti-cancer treatment, research efforts have been underway to develop innovative pharmaceutical systems to scientifically evaluate novel IO agents. Syngeneic mouse allografts are currently the standard *in vivo* tumor model for IO drug development, contributing to proof of efficacy, first in human (FIH) dose projections, safety evaluations, as well as clues about drug mechanism of action (MOA) and biomarkers. Syngeneic mouse tumor models display mouse immune responses, and lack human diversity for translational use (2). Efforts have been underway to humanize mouse immune systems, and consequently, to expand the diversity of available models (3). Transgenic models have also been developed to harbor genetic lesions in key driver oncogenes, and spontaneously yield tumors in immune competent mice (4–5). Both humanized and transgenic mouse tumor models are as of yet time consuming and expensive for widespread use in IO drug development.

1.1 IO Drug Combination Development

IO drug combination drug assessment is among the many strategies currently being investigated to enhance IO efficacy (6). Pre-clinical identification of candidate drug combinations remains perplexing at many levels. To begin with, the term 'synergy' with respect to drug combinations has many different scientific definitions (7). Preclinical *in vivo* investigations are conducted in: (*i*) a small collection of *replicated* tumors, (*ii*) from a limited population of different syngeneic mouse models, (*iii*) treated in multiple ways (8). Further, syngeneic mouse models are selected for experimentation based on experience and/or available biological profiling. A suitable mouse to scientifically evaluate a novel drug combination may not exist in some cases. By contrast, clinical treatment outcomes are summarized over many different patients, treated similarly by study arm (9). While a favorable drug interaction could in theory be inferred clinically, it might not be a prerequisite for regulatory approval depending on unmet medical needs. Moreover, pharmacological investigations around combination drug doses and schedules differ with respect to feasibility and stage (10).

1.2 IO Drug Combination Evaluation

Our pre-clinical null hypothesis is

H_0: the drug combination is not superior to monotherapies

versus the experimental hypothesis

H_1: the drug combination is superior to monotherapies

in a study design of n independent replicated mouse tumors implanted in a single type of syngeneic mouse model, randomized to four treatment arms: control, drug *A*, drug *B*, and *A*+*B* (7). In this experimental set up, we require investigations around combined dosing and safety as having been performed, our aim not being to optimize drug development, rather to evaluate efficacy. While we acknowledge that drug synergies and combined dosing are related, we treat pharmacological explorations and efficacy evaluations separately, to arrive at valid scientific conclusions. Further, while the data between different types of syngeneic mouse models could be analyzed jointly, depending on available biological profiling suspected biases could be introduced. As a counter example, a scientific criteria might require proof of efficacy in one type of mouse model and not another, i.e., relying on a biological premise.

Some common endpoints in murine tumor experiments include the tumor free response (*TFR*) rate, tumor growth delay (*TGD*) or log cell kill (*LCK*), tumor growth inhibition (*TGI*), progression free and overall survival (*PFS*, *OS*) (11). An important consideration when selecting an endpoint is that syngeneic mouse tumors grown in inbred mouse strains have been observed to display *categorically* variable responses to the same IO therapeutic intervention, whereby subgroups show different outcomes: resistance, partial, or complete response (see also (8,12)). In Figure 1(a) n=8 longitudinal MC38 tumor volumes mm^3 are shown from mice treated with an isotype control antibody and in Figure 1(b) treated with an *αCtla4*(*CTLA4* human, *Ctla4* mouse) antibody. Log-linear tumor growth was observed in the control arm. In Figure 1(b) 2 tumors (black) showed little-to-no response (*NR*) to *αCtla4*, 4 mice (red) experienced transient tumor regression followed by progression/re-growth for a *TGD* response (*DR*), and 2 mice (purple) were *TF* though end of follow up.

Statistical comparisons of treatments in mice are therefore complicated by the various ways that drug combinations can be superior to single agents. Analysis of the overall trend in just one endpoint might be sufficient for making unbiased comparisons if mice respond categorically the same to a treatment, e.g., when all mice have *DR*'s to treatment, and otherwise can under or over-estimate a single endpoint if outcomes vary categorically, while not necessarily simplifying the analysis in an efficient way. We introduce Bayesian analysis in methods evaluating combination drug superiority, but first we introduce some basic Bayesian analysis concepts.

1.3 Introduction to Bayesian Analysis of a *TFR* Rate

Consider the *TFR* rate associated with a drug in repeated independent, randomized, and uniform experimentation of replicated tumors from a single mouse model, denoted θ_{TF}, which is unknown and unobservable in the population being the true *TF* frequency of treated animals in endless experimentation. Before collecting data about the θ_{TF}, the likelihood of the count y of *TF* events in n trials is specified as $y \sim Binomial(n, \theta_{\text{TF}})$. Bayesian analysis begins by specifying prior *probabilistic* beliefs about θ_{TF} through the probability function $Pr(\theta_{TF})$, or 'prior'. Suppose *a priori* that θ_{TF} was assumed equally likely to be any value between 0 and 1, following a $Beta(1, 1)$ distribution, with prior uncertainty reflected in Figure 2(a). After observing outcomes from a designed experiment, the posterior probability function $Pr(\theta_{TF} \mid data)$ is computed by Bayes Theorem

$$Pr(\theta_{\text{TF}} \mid data) \propto Pr(\text{ data} \mid \theta_{\text{TF}}) \cdot Pr(\theta_{\text{TF}}), \qquad (1)$$

by updating $Pr(\theta_{TF})$ with data. $Pr(\theta_{TF} \mid data)$ tells us what we have learned about θ_{TF} given empirical observations since the outset. In this example, the posterior is derived in closed form as

$$\theta_{TF} \mid y, n \sim Beta(1+ y, 1+ n - y). \qquad (2)$$

For illustration, suppose interest was in evaluating the experimental hypothesis H_1: $\theta_{TF} \geq {}^1/_2$ in a study of an anti-tumor treatment in n=8 mice, and that by study end the treatment yielded ${}^5/_8$ *TF* mice. Before observing outcomes to treatment, the prior probability of H_1: $\theta_{TF} \geq {}^1/_2$ was $Pr(\theta_{TF} \geq {}^1/_2) = 50\%$ and after observing outcomes the posterior probability $Pr(\theta_{TF} \geq {}^1/_2 \mid {}^5/_8$ mice TF) = 75%. The resulting posterior mean of θ_{TF} equals 0.60. An 80% equal tail credible interval for θ_{TF} is (0.40, 0.79). The values 0.40 and 0.79 are depicted in Figure 2(b), to illustrate the 80% posterior density region with equal probability tails. In Figure 2(c) the posterior 80% credible interval was narrower, if we assumed instead that the replicate sample size had been larger, instead, observing ${}^{10}/_{16}$ mice TF. In Figure 2(d), the distribution of the non-TFR, or $1-\theta_{\text{TF}}$, is shown; a reflection of the posterior density of θ_{TF}. Note that wider credible intervals can be found by allowing the tail probabilities to vary from one another in this example, by applying a Highest Posterior Density Region (HPD) region calculation.

2. Methods

2.1 Bayesian Semi-Mixture Response Model

The log tumor volume y_{tij} at time t for mouse tumor replicate $j=1,\ldots,n_{\text{i}}$ administered treatment i was modeled as

$$y_{\text{tij}} = \beta_{0\text{ij}} + \beta_{1\text{ij}} t - \beta_{2\text{ij}} \cdot g(t; \vartheta_{\text{ij}}) + \varepsilon_{\text{tij}} \qquad (3)$$

where β_{0ij} and β_{1ij} are the intercept and slope of the log-linear tumor growth component of the model, the term $\beta_{2ij} \cdot g(t;\vartheta_{ij})$ accounts for treatment associated decay or tumor growth inhibition, and the term ε_{tij} accounts for unexplained residual noise. The function $g(t;\vartheta_{ij})$ was assumed to be bounded below and monotonically increasing, asymptotically in time to an upper limit. By definition $\beta_{21\text{j}} = 0$ in the control arm, and otherwise, $\beta_{2\text{ij}}$ was modeled as a mixture: $\beta_{2\text{ij}} = 0$ indicating no tumor inhibition effect and $\beta_{2\text{ij}} > 0$ indicating tumor regression followed by tumor progression. We specified $g(t|\,\vartheta)$ as a Gaussian cumulative distribution function, with mean and standard deviation $\vartheta = (\mu, \nu)$. Conveniently, μ defined the tumor growth decay inflection point and ν the shape or duration. In our case studies our choice of g performed well. Scientific evidence to select g would be helpful, if available. Finally, ε_{tij}'s were modeled as independent Gaussian random noise with mean 0 and standard deviation σ.

Mixture modeling can be helpful when a categorical treatment outcome is uncertain or multiple outcomes can plausibly be interpreted from variable data. By modeling a probability between 0 and 1 of an uncertain treatment outcome instead of attempting a 0/1 prediction, uncertainty can be holistically and transparently modeled rather than dealt with in an ad-hoc and possibly biased way. Our semi-mixture approach, modeling probabilities of: *TFR*, *DR*, or *NR*, is introduced as follows. If mouse *j*'s tumor regressed below 100mm^3 after receiving treatment *i* and did not re-emerge by study end we observed mouse *j TF*, and fixed mouse *j*'s *TF* probability $\theta_{TF}^{(ij)} = 1$. Tumor volumes of *TF* mice were not modeled; the *TF* outcome being *observed* and volumes providing nothing additional for our analysis, though possibly of interest for other purposes, e.g., time to eradication. Otherwise, the *TF* probability was set to 0, in which case the posterior probability $\theta_{DR}^{(ij)}$ of *DR* was estimated by Bayesian modeling of longitudinal tumor volumes. The posterior magnitude of treatment durability, i.e., fitted non-linear tumor regression and re-growth, was computed simultaneously with $\theta_{DR}^{(ij)} = P(\,\beta_{2ij} > 0\,|\text{data})$ by Markov Chain Monte Carlo Reversible Jump, as otherwise mouse tumor *j* was estimated to be unresponsive to treatment (*NR*) and growing log-linearly with probability $\theta_{NR}^{(ij)} = 1 - \theta_{DR}^{(ij)}$.

We specified the prior distributions

$$\beta_{1\text{ij}} \sim \text{TruncNorm}(4.5, 0.20; 0, \infty) \quad (4)$$

$$\beta_{2\text{ij}} \sim \text{TruncNorm}(0.10, 0.01; 0, \infty) \quad (5)$$

$$\beta_{2\text{ij}} \sim (1 - \xi_{ij}) \cdot 1_{\{0\}} + \xi_{ij} \cdot \text{TruncNorm}(3, 0.50; 0, \infty) \quad (6)$$

$$\xi_{ij} \sim Bernoulli(0.50) \quad (7)$$

$$\mu \sim \text{N}(10\,, 5) \quad (8)$$

$$\nu^{-2} \sim Gamma(50, 100) \quad (9)$$

$$\sigma^{-2} \sim Gamma(10, 0.10), \quad (10)$$

where ξ_{ij} is the mixture indicator, equaling 1 for a *DR* and 0 for *NR*. These priors were chosen by experience and historically available MC38 mouse tumor data. The model was programmed and fit in the R language, by Markov Chain Monte Carlo (MCMC) Reversible Jump, for 10,000 iterations, discarding the run in of 50 iterations, and thinning by 10 iterations. No borrowing of information was allowed between tumors or arms. MCMC outcome counts, i.e., posterior random counts denoted *#NR* or *#DR* at each MCMC iteration, as well as the given or observed *#TF* count, were all input into a *Dirichlet*(1+*#NR*,1+*#DR*,1+*#TF*) distribution, to generate response rate aggregates in each treatment arm.

2.2 Conditional Log Cell Kill

We introduce what we call the conditional *LCK* (*cLCK*) endpoint, a measure of treatment durability conditioned on a *DR* outcome. *LCK* summarizes the distribution of *TGD* over all treated tumors, normalized by tumor volume doubling time in the control arm, *naïve* to disparate outcomes (12). *LCK* as a global measure of efficacy can be misleading. For instance, two treatments can yield similar *LCK*, though have different *NR*, *DR*, and *TF* outcome frequencies. Differences in *LCK* may not be reflective of differences in treatment durability alone and influenced by *NR's* and *TFR*'s to treatment. Precaution should be taken, even when relying on censoring adjustments.

Since we do not observe *DR* with complete certainty, we measure *cLCK* as a weighted average of the fitted posterior *LCK* for each mouse tumor j =1,.., n_i weighted by $\theta_{DR}^{(ij)}$, which is 0 for eradicated tumors and small for tumors with nascent growth inhibition. In our posterior computation, we sampled *cLCK* by treatment arms i = 2, …, as

$$cLCK_i^{(s)} \sim \frac{\Sigma_j(\beta_{2ij}^{(s)} \div \log(10) \left| \beta_{2ij}^{(s)} > 0 \right.)}{\Sigma_j I(\beta_{2ij}^{(s)} > 0)} \qquad (10)$$

given posterior samples $\beta_{2ij}^{(s)}$ (11). Our approach requires three assumptions, that the treated tumors: (*i*) achieved overall log-linear progression/re-growth, (*ii*) progressed/re-grew parallel to the control arm, and that (*iii*) tumor progression was not too wavy. Censoring adjustments are not necessary, or specification of a target tumor size, if these conditions are satisfied.

2.3 Superiority in Dual Endpoints

Comparing the drug combination *A*+*B* to drug *A* and drug *B* can be complicated with dual endpoints by the multiple ways *A*+*B* can be superior. To illustrate, let drugs *A* and *B* have different trade-offs, drug *A* with a 20%*TFR* and 0.75*cLCK* and *B* with a 40%*TFR* and 0.50*cLCK* shown in Figure 3. *A*+*B* is superior if it yields better durability *and* activity to drugs *A* and *B* alone. In this example, superiority calls for >40%*TFR* and >0.75*cLCK*. An alternative to superiority with dual endpoints is an improvement in just one, an event we term partial superiority. Bayesian evaluation is performed by assigning a probability to the statement "superiority." A statistical finding is made by comparing *Pr*("superiority" | data) > *C*, a threshold to control error. If scientific and other considerations warranted, the drug combination utility criteria could be defined in various other ways, e.g., non-linear / asymmetric trade-offs or decision trees, applicable in a statistical framework and not discussed further here.

3. Results

3.1 αCTLA4 Retrospective Case Study

We conducted a retrospective Bayesian investigation of data from a study combining *αCtla4* with a small molecule immuno-modulator. The original study included each of *n*=8 MC38 mice treated with: isotype control antibody, *αCtla4*, a small molecule immuno-modulator, or the combination of *αCtla4* and the immuno-modulator. Figure 1(c) displays longitudinal tumor outcomes to *αCtla4* combined with a small molecule immuno-modulator. Treatment responses were lacking for the small molecule alone while activity was observed for the combination. Fitted tumor volumes and rates of *TFR*, *DR*, and *NR* are displayed in Figure 4. For *αCtla4* the posterior distribution of the

NR rate (black) had a skewed long right tail with a mode near 0.20. The posterior distributions of the *DR* (red) and *TF* (purple) rates were unimodal with modes not surprisingly near 0.50 and 0.25.

Our primary test hypothesis was drug combination superiority to single agents in the *TFR* and *cLCK* endpoints. Since Bayesian analysis does not utilize a null distribution, we determined the threshold *C* for making a statistical conclusion of superiority at the 10% significance level, see Table 1 for operating characteristics. Our null scenario was that the small molecule alone was inactive while single agent α*Ctla4* achieved 25%TFR and 0.75cLCK. We conducted *S*=250 simulated data runs for each scenario. Simulation parameters were specified as

$$\beta_{1ij} \sim \text{TruncNorm}(4.5, 0.10; 0, \infty) \quad (11)$$

$$\beta_{2ij} \sim \text{TruncNorm}(0.12, 0.01; 0, \infty) \quad (12)$$

$$\beta_{2ij} \sim (1 - \xi_{ij}) \cdot 1_{\{0\}} + \xi_{ij} \cdot \text{TruncNorm}(2.3\text{cLCK}, 0.576\text{cLCK}; 0, \infty) \quad (13)$$

$$\xi_{ij} \sim Bernoulli(\theta_{DR}) \quad (14)$$

$$\mu_{i1} \sim \text{N}(10\ , 1) \quad (15)$$

$$\nu_{i2}^{-2} \sim Gamma(10, 100) \quad (16)$$

$$\sigma^2 = 0.2 \quad (17)$$

where *cLCK*, θ_{DR}, and θ_{TF} were user defined and allowed to vary between scenarios, in Table 1. The threshold C = 45% provided false positive error control ≤ 10% and reasonable power for alternatives listed in Table 1. For partial superiority a higher threshold C = 80% was required to control the false positive rate for either endpoint, also at ≤ 10%.

The Bayesian posterior probability of superiority, computed as

$$Pr(cLCK_{A+B} > cLCK_A \cap cLCK_{A+B} > cLCK_B \cap TFR_{A+B} > TFR_A \cap TFR_{A+B} > TFR_B \mid \text{data}) \quad (18)$$

was 36%, and we concluded superiority was not supported by the data, conditionally in retrospect. Joint posterior probability contours are shown in Figure 1(d) for the *TFR* rate and *cLCK* radiating in descending probability for each treatment arm. There was a 73% posterior probability that the drug combination yielded a superior *TFR*, while the posterior probability of *cLCK* improvement alone was 42%. We concluded the combination not partially superior conditionally in retrospect at the 10% level. An obvious criticism of this sophisticated approach is the small *n*. The alternative scenarios listed in Table 1 were large and meaningful in magnitude. We acknowledge attention to statistical considerations in Discussion.

3.2 Immune Agonist Combination Retrospective Case Study

Immune agonist antibodies denoted *A* and *B* were administered alone or in combination in a study conducted of *n*=8 MC38 tumor bearing mice in each of four treatment arms, including a control arm, Figure 5. We retrospectively tested the immune agonist combination for superiority. Visually, mild-to-no treatment responses were seen for *A*, some treatment responsiveness was seen to *B*, and the combination yielded 2 *TF* mice with some treatment durability, Figure 5.

In contrast to the previous case study, we tested the drug combination superiority hypothesis at the 1% false positive significance level, and partial superiority at the 5% level. These significance levels reflected desire for greater conviction that if statistically significant, the findings would have a lower chance of being false positives. Investigation of the threshold for making statistical conclusions was performed similar to

the previous case study. Our null scenario was that agonist *A* was inactive, while single agent *B* achieved a true 10%*TFR* and 0.25*cLCK*. The operating characteristics are listed in Table 2. The posterior probability cut-off *C*=30% was identified to protect the level for the hypothesis test of combination superiority at 1%, while providing reasonable power for alternatives. The respective posterior probability cut-offs were identified to protect test levels for partial superiority: *C*=80% for *TFR* and *C*=45% for *cLCK*, each at 5%.

Joint posterior probability contours are shown in Figure 6 for the *TFR* on the *x*-axis and *cLCK* on the *y*-axis radiating out as descending probabilities for each treatment arm: *A* (grey), *B* (tan), and A+B (brick-red). The joint posterior probability density for *A* is concentrated in the lower left corner of Figure 3, with mean (8%*TFR*, 0.12*cLCK*). Some growth delay was seen for *B*, joint posterior mean (16%*TFR*, 0.28*cLCK*). The joint posterior mean of *A*+*B* was (31%*TFR*, 0.61*cLCK*) and the posterior probability of superiority was found to be 78%. We concluded *A*+*B* superior to *A* and *B* in activity and durability, conditionally in retrospect, at the 1% level.

4. Discussion

We developed a Bayesian application to evaluate IO drug combinations in mouse studies, in dual endpoints for activity and durability. Categorically heterogeneous outcomes to treatment were seen in our case studies. In our first case study the drug combination was concluded not to be superior or partially superior to single agents, though some activity was observed for the combination over *αCTLA4* alone. In our next case study, we found the combination of immune agonists statistically superior in both activity and durability. Inference in dual endpoints can provide insights into qualitatively distinct benefits of IO combinations. In our case studies, note that the posterior distributions of the endpoints were not normally distributed. While Bayesian findings do not depend on a null distribution or rely on normality, we highly recommend that pre-clinical investigators thoughtfully consider sample size and risk when investigating drug combinations, particularly with prior information. Experimental design with prior information is the subject of on-going work, along with robust extensions of our modeling framework.

Anti-cancer immunotherapeutic drug combination development is progressing rapidly. This is a revolutionary time for cancer care, as greater clinical benefits are anticipated by combining drugs with multiple modes of action. Inherent challenges include driving promising drugs and combinations to the forefront as quickly and efficiently as possible.

References

1. Zugazagoitia J, Guedes C, Ponce S, Ferrer I, Molina-Pinelo S, Paz-Ares L. Current Challenges in Cancer Treatment. Clin Ther 2016;38(7):1551–66.
2. Kuzu OF, Nguyen FD, Noory MA, Sharma A. Current state of animal (mouse) modeling in melanoma research cancer growth and metastasis. Cancer Growth and Metastasis 2015;8:81–94.
3 . Holzapfel BM, Thibaudeau L, Hesami P, Taubenberger A, Holzapfel NP, Mayer-Wagner S, Power C, Clements J, Russell P, Hutmacher DW. Humanised xenograft models of bone metastasis revisited: novel insights into species-specific mechanisms of cancer cell osteotropism. Cancer Metastasis Rev 2013;32:129–45.
4. Singh P, Schimenti JC, Bolcun-Filas E. A mouse geneticist's practical guide to crispr applications. Genetics 2015;199:1–15.

5. Champiat S, Ferté C, Lebel-Binay S, Eggermont A, Soria JC. Bridging mutational load and immune checkpoints efficacy. Oncoimmunology 2014;3:e27817.
6. Carlino MS, Long GV. Ipilimumab Combined with Nivolumab: A Standard of Care for the Treatment of Advanced Melanoma? Clin Cancer Res 2016; clincanres.2944.2016.
7. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm236669.pdf
8. Houghton PJ, Morton CL, Gorlick R, Lock RB, Carol H, Reynolds CP, Kang MH, Maris JM, Keir ST, Kolb EA, Wu J, Wozniak AW, Billups CA, Rubinstein L, Smith MA. Stage 2 combination testing of rapamycin with cytotoxic agents by the pediatric preclinical testing program. Mol Cancer Ther 2010;9(1):101–12.
9. Wei L. An efficient design for a study comparing two drugs, their combination and placebo. Stat in Med 2006; 25(12):2043–58.
10. Chou TC. Drug combination studies and their synergy quantification using chao-talalay method. Cancer Res 2010;70(2):440–46.
11. Wu J. Assessment of antitumor activity for tumor xenograft studies using exponential growth models. Journ of Biopharm Stat 2001;21:472–83.
12. Laajala TD, Corander J, Saarinen NM, Makel K, Savolainen S, Suominen MI, Alhoniemi E, Makel S, Poutanen M, Aittokallio T. Improved statistical modeling of tumor growth and treatment effect in preclinical animal studies with highly heterogeneous responses in vivo. Clin Cancer Res 2014;18(16):4385–96.
13. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB. Mixture Models. In: Bayesian data analysis, Second Edition, CRC Press; 2013. p. 463-80.
14. Ribba B, Holford NH, Magni P, Trocóniz I, Gueorguieva I, Girard P, Sarr C, Elishmereni M, Le Friberg CK. A review of mixed-effects models of tumor growth and effects of anticancer drug treatment used in population analysis. Cpt Pharm Syst Pharmacol 2014;3:e113.
15. Zhao L, Morgan MA, Parsels LA, Maybaum J, Lawrence TS, Normolle D. Bayesian hierarchical changepoint methods in modeling the tumor growth profiles in xenograft experiments. Clin Cancer Res 2010;17(5):1057–64.
16. Jensen SM, Pipper CB, Ritz C. Evaluation of multi-outcome longitudinal studies. Stat Med 2015;34(12):1993–2003.

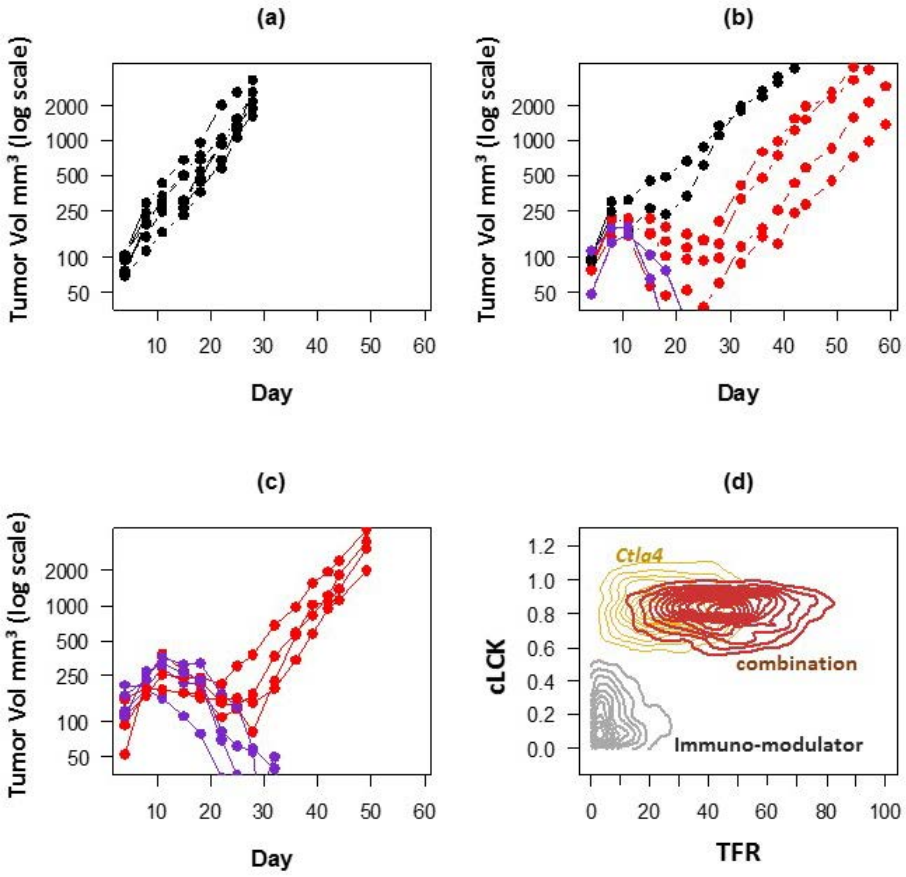

Figure 1: (a) Isotype control treated tumors, (b) tumor volumes for tumors treated with single agent *αCtla4* or (c) treated with *αCtla4* combined with an immuno-modulator, (d) the joint posterior probabilities shown as radiating contours, i.e., descending probabilities, comparing *αCtla*4 (tan), small molecule immuno-modulator (grey), and the combination (brick-red).

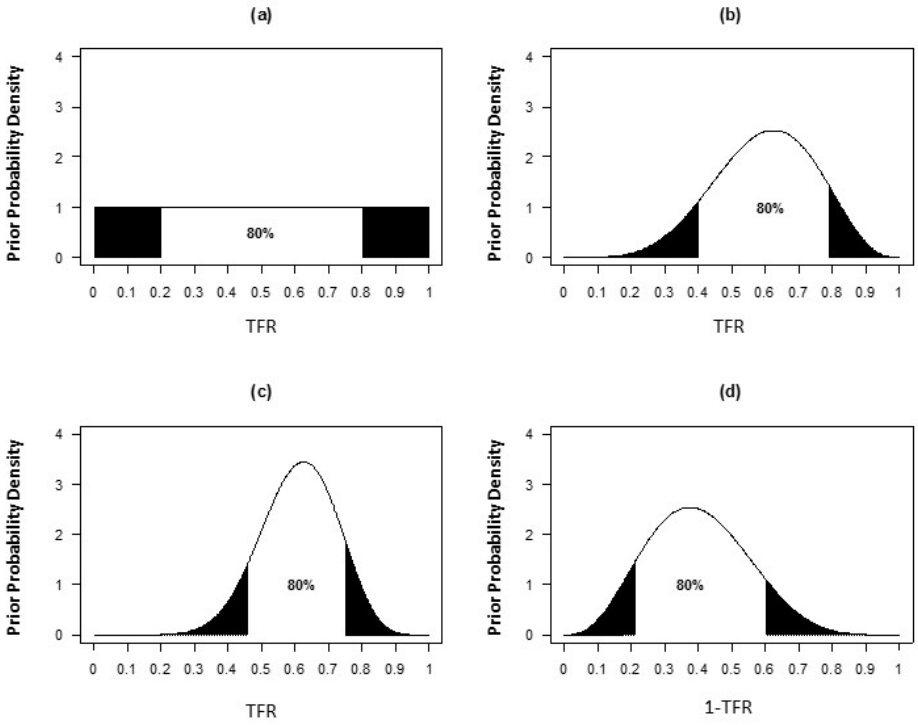

Figure 2: Illustration of Bayesian analysis about the TFR rate with hypothetical outcomes: (a) prior to observing data, probabilistic beliefs about the TFR rate are flat over the interval (0, 1), (b) after observing $^{5}/_{8}$ TF mice, the posterior density of the TFR rate is updated to have a posterior mode near 60% and heavy tails, (c) increasing the sample size to $^{10}/_{16}$ mice TF, the posterior distribution is more concentrated around 60%. (d) posterior density of the non-TFR (1-TRF) rate, reflection of (b).

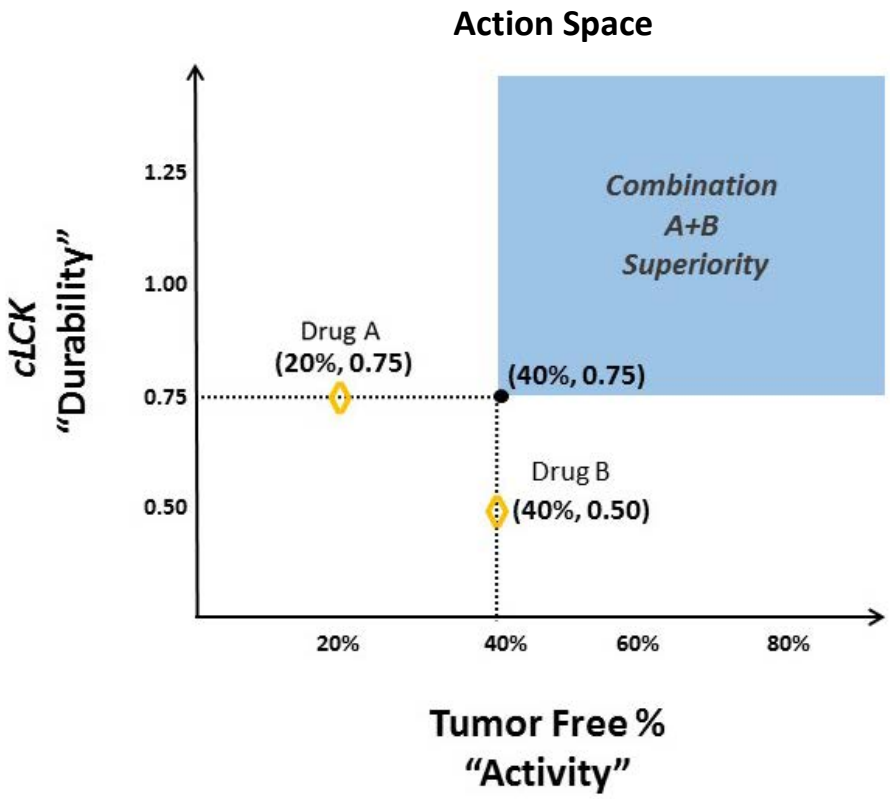

Figure 3: Illustration of dual endpoint criteria in action space: drugs *A* and *B* have different tradeoffs, drug *A* with a 20%*TFR* and 0.75*cLCK* and *B* with a 40%*TFR* and 0.50*cLCK*. *A+B* is superior if it yields better durability *and* activity to drugs *A* and *B* alone. The point (40%, 0.75) is marked at the superiority boundary.

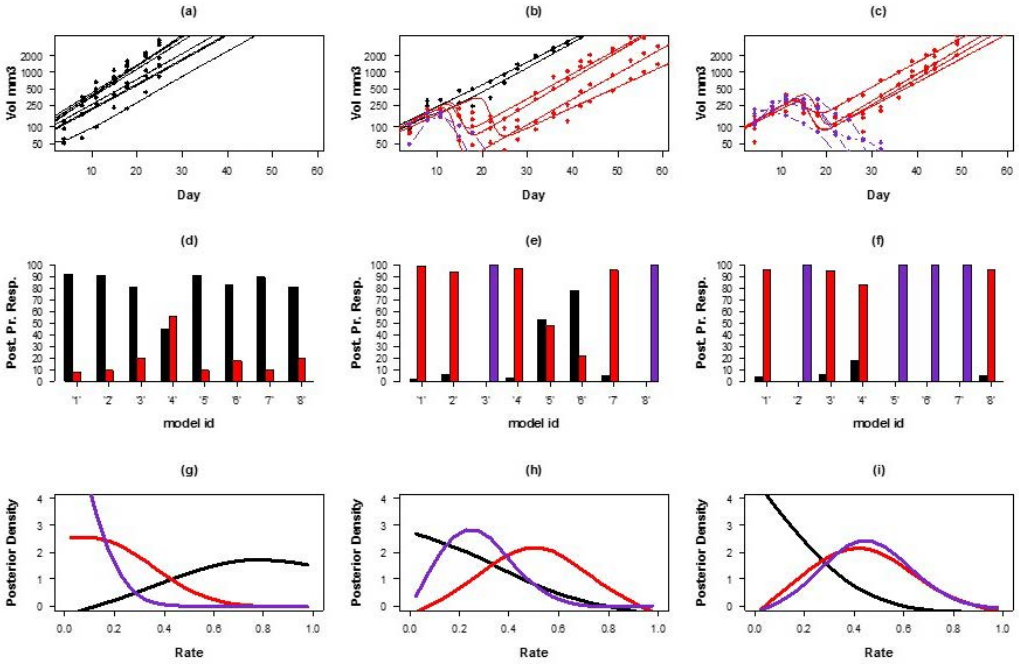

Figure 4: Fitted posterior means of tumor volumes: (a) *αCtla4*, (b) immuno-modulator, (c) *αCtla4*+immuno-modulator, (middle-row) fitted tumor-wise posterior probabilities for *NR*(black), *DR*(red),*TF*(purple): (d) *αCtla4*, (e) immuno-modulator, (f) *αCtla4*+immuno-modulator, and (bottom row) arm-wise aggregated posterior densities of response rates *NR*(black), *DR*(red),*TF*(purple): (g) *αCtla4*, (h) immuno-modulator, (i) *αCtla4*+immuno-modulator.

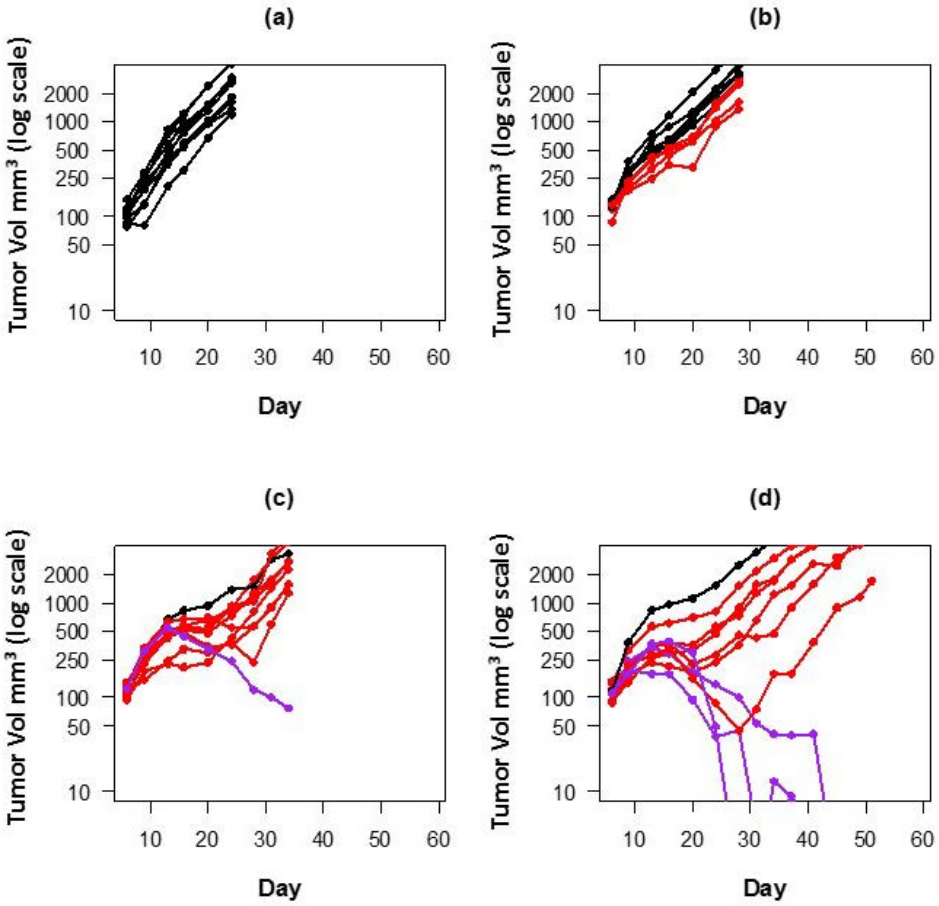

Figure 5: Drug combination comparison study, tumor volumes in treatment arms: (a) Control, (b) Immune Agonist A (c) Immune Agonist B, (d) Combination A+B.

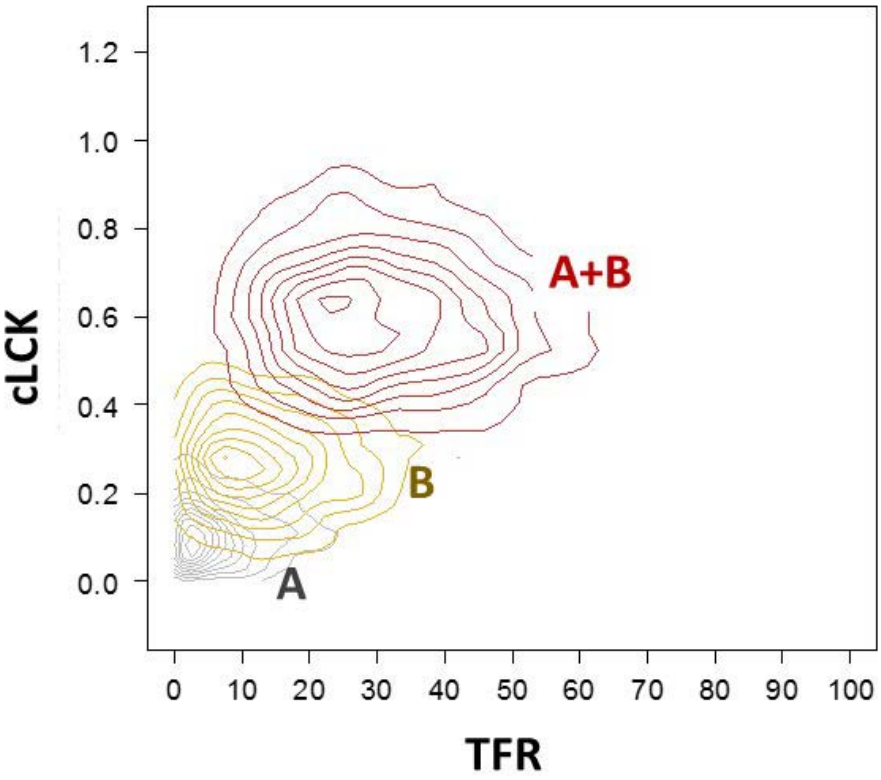

Figure 6: Drug combination comparison study. Joint posterior probabilities *cLCK* and *TFR* rate shown as radiating contours for immune agonists A, B, and A+B.

Table 1: Detection rates: superiority with belief threshold C = 45% and partial superiority, *C* = 80%, per 250 simulated data sets.

*Scenario**	*Test Hyp.**	*TFR*	*cLCK*	*Detection Rate*
null/base	PS	0.25	0.75	10%
null/base	S	0.25	0.75	7%
PS	PS	0.25	1.25	88%
PS	PS	0.25	1.5	93%
PS	PS	0.50	0.75	49%
S	S	0.50	1.25	74%
S	S	0.50	1.50	68%
PS	PS	0.75	0.75	86%
S	S	0.75	1.25	92%
S	S	0.75	1.50	98%

*PS: Partial Superiority, S: Superiority

Table 2: Detection rates: superiority with belief threshold *C*=30%, and partial superiority, *C* = 80% *TFR* and *C* = 45% *cLCK*, per 250 simulated data sets.

*Scenario**	*Test Hyp.**	*TFR*	*cLCK*	*Detection Rate*
null/base	PS	0.10	0.25	5%
null/base	S	0.10	0.25	1%
PS	PS	0.10	0.50	79%
PS	PS	0.10	0.75	96%
PS	PS	0.25	0.25	31%
S	S	0.25	0.50	50%
S	S	0.25	0.75	62%
PS	PS	0.50	0.25	75%
S	S	0.50	0.50	89%
S	S	0.50	0.75	95%

*PS: Partial Superiority, S: Superiority