SynBa: Improved estimation of drug combination synergies with uncertainty quantification

Haoting Zhang,[1,2] Carl Henrik Ek,[1] Magnus Rattray[3, 4] and Marta Milo[5]

[1]Department of Computer Science and Technology, University of Cambridge, Cambridge, UK, [2]Health Data Research UK, London, UK, [3]Division of Informatics, Imaging and Data Sciences, University of Manchester, Manchester, UK, [4]Alan Turing Institute, London, UK and [5]Oncology Data Science, Oncology R&D AstraZeneca, Cambridge, UK

Abstract

There exists a range of different quantification frameworks to estimate the synergistic effect of drug combinations. The diversity and disagreement in estimates make it challenging to determine which combinations from a large drug screening should be proceeded with. Furthermore, the lack of accurate uncertainty quantification for those estimates precludes the choice of optimal drug combinations based on the most favourable synergistic effect. In this work, we propose SynBa, a flexible Bayesian approach to estimate the uncertainty of the synergistic efficacy and potency of drug combinations, so that actionable decisions can be derived from the model outputs. The actionability is enabled by incorporating the Hill equation into SynBa, so that the parameters representing the potency and the efficacy can be preserved. Existing knowledge may be conveniently inserted due to the flexibility of the prior, as shown by the empirical Beta prior defined for the normalised maximal inhibition. Through experiments on large combination screenings and comparison against benchmark methods, we show that SynBa provides improved accuracy of dose-response predictions and better-calibrated uncertainty estimation for the parameters and the predictions.

Introduction

With the increased use of small molecule drugs in monotherapy and drug combination treatments, off-target toxicity and resistance to treatments are becoming clear challenges. The use of combinations of drugs offers a possible solution to reduce toxicity minimising doses and bypassing resistance with alternative targeting. Thanks to the development of high-throughput approaches to screen drug-sensitivity in cell lines, dose-response data for a large number of combinations have been made available. Examples include the AstraZeneca-Sanger DREAM challenge (DREAM) [1], NCI-ALMANAC [2], DrugComb [3], and the screening data from the Wellcome Sanger Institute [4]. The availability of these datasets makes it possible to predict the effect of drug combinations from modelling drug-sensitivity data, leveraging information from biological features of the cell lines and chemical characteristics of the drugs [5].

To understand how synergistically drugs can work when combined, models with a quantification framework are required. Traditional quantification frameworks usually involve a null surface, based on a number of set assumptions. The Bliss model is based on the Multiplicative Survival Principle [6], whereas the Loewe model is based on the Dose Equivalence Principle [7]. In the past decade, parametric methods have started to emerge as alternatives to the above. These include MuSyC [8, 9], BRAID [10] and the Effective Dose model [11]. All these frameworks are based on assumptions and parameterisations different from each other and as a result of this, the outputs from these models often disagree. Although there have been efforts to unify the frameworks, this is still an open problem for the field.

To define a common framework for drug combination, we need a less ambiguous definition of *synergy* [12]. When a combination is said to be *synergistic*, it is unclear whether it implies that the combination is desirable in terms of its *potency* or its *efficacy*. Potency is the amount of dosage required for a drug to produce a specified effect, whereas efficacy is the degree of the beneficial effect produced by the drug [8]. A strong synergistic potency implies the toxicity may be reduced when the drugs are combined, which is crucial for avoiding overdose. On the other hand, a strong synergistic efficacy implies that the combination increases the maximal possible effect. Both aspects are relevant for progressing with further pre-clinical and subsequent clinical investigation. An ideal combination would be potent and effective. However, in clinical research there are situations where a drug partner is requested only for enhancement of potency and not to increase efficacy. Until recently, potency and efficacy are entangled within the concept of *synergy* as defined in the traditional quantification frameworks. To tackle this problem, Meyer et al. [8] developed MuSyC, a framework that decouples potency and efficacy following the principles of the generalised Hill equation. Although the MuSyC approach is effective for modelling both potency and efficacy, the model does not fully explore the challenge of properly handling uncertainty using model-based estimation. There are various sources of uncertainty associated with drug-sensitivity modelling. Firstly, the biology of dose-response relationships is still unknown despite existing efforts. This leads to uncertainty associated with insufficient scientific knowledge. Secondly, there are systematic and random errors arising from the experimental procedures. Thirdly, biological variation exists among cell lines of the same type, resulting in a further source of noise. Finally, uncertainty also stems from limited information that can be extracted due to the small size of available data (*epistemic uncertainty*).

Given these multiple sources of uncertainty, it is often impossible to reach a true estimate of the quantities of interest, e.g potency of a monotherapy, or the synergistic effect (in terms of either potency or efficacy) of a combination. Since the *best* single deterministic estimates might not be reached, here we focus on accurately quantifying uncertainty in the model (e.g. the credible range of model parameter estimates/predictions) to quantify a devitiaon from the true estimate given the noise.

The majority of the existing frameworks for drug combinations either do not compute the uncertainty or

estimate it by standard error [11] or parametric bootstrap [9], each of which contains its own unrealistic assumptions. The standard error is only accurate when a large number of samples is available, which is not the case for most pharmacology datasets. Similarly, for the parametric bootstrap, an accurate uncertainty estimation relies on a sufficiently large number of observations to be reasonably close to the truth. In either case, the uncertainty estimation will often be inaccurate due to the small data sizes. For this reason, here we approach uncertainty estimation with a Bayesian framework, which incorporates the uncertainty by treating all parameters of interest as probabilistic quantities. This enables us to continuously model the uncertainty in our estimates as the number of measurements grows, without becoming over-confident. Moreover, the estimated quantities of interest are obtained simultaneously with their uncertainties, making this approach computationally efficient.

In the current literature, there are examples that use a probabilistic model to incorporate uncertainty in their outputs. Shapovalova et al. developed Hand-GP [13], a non-parametric model based on the combination of the Hand model with Gaussian processes, providing more believable uncertainty estimation than MuSyC in some cases. However, Hand-GP does not incorporate the 1D Hill equation that imposed biological constraints useful for providing interpretability of the model outputs. For example, monotonicity of the monotherapy fitted curve is not enforced in Hand-GP, meaning that it is possible that the model produces a dose-response surface that is physically impossible. More importantly, due to the non-parametric structure of Hand-GP, the parameters describing potency and efficacy are lost. Ronneberg et al. developed `bayesynergy` [14], a Bayesian framework that models synergistic interaction effects using Gaussian Processes, which provide uncertainty quantification. Although flexible, its formulation is still based on the Bliss independence assumption that is biased against combinations of drugs with a moderate level of efficacy, compared to those with strong or weak efficacy [9]. `bayesynergy` is also not able to separate out synergistic potency and efficacy.

Here, we design a flexible Bayesian framework to infer Synergistic effects of drug Combination (**SynBa**) where (1) the classic Hill equation is preserved to produce estimates of efficacy and potency (2) the existing biological knowledge or insight from historical data may be conveniently added through the prior distribution over parameters in the model. In SynBa we will use MuSyC as a baseline framework to decouple synergistic potency and efficacy and add probabilistic inference to provide outputs and their associated uncertainty. This is to design a framework estimating the most favourable scores and efficiently provide optimal candidates to the drug discovery pipeline with actionable decision-making criteria that can be easily derived from the model outputs.

With our approach, when a combination is predicted to be synergistic, a level of confidence will be quantified for this prediction, as well as whether the synergy stems from an improvement in efficacy or in potency. If the synergy prediction has high associated uncertainty, then further laboratory experiments may be required before proceeding to the next step in the pipeline. The ability to quantify all these variables and their clear interpretation will provide *actionable metrics* for the subsequent stages of the drug discovery pipeline, which is an unmet need. Our model offers the ability to define any consequential decision-making step, whether pre-clinical or clinical, based on accurate and calibrated uncertainty quantification.

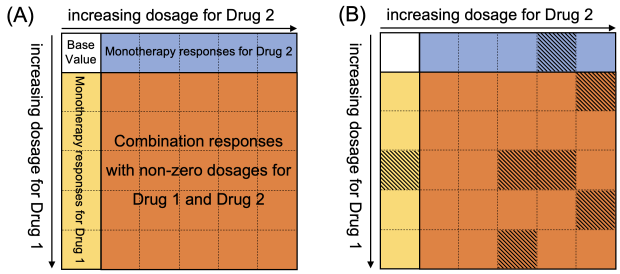

Fig. 1. Description of the dose-response datasets for the proposed framework. (A): A typical dose-response matrix, where the top-left entry is the base value, the first row contains the monotherapy responses for the first agent, the first column contains the monotherapy responses for the second agent. The remaining entries are the responses when a combination of the two drugs is applied. (B): A dose-response matrix where not all responses are available for training. The shaded cells represent the test data.

Methods

Our proposed method can be used for both analysing existing dose-response data and predicting unseen dose-response data for a given monotherapy $\mathcal{D} = (X, Y) = \{(x_i, y_i) \mid i = 1, ..., N\}$ or a given combination $\mathcal{D} = (X, Y) = \{(\mathbf{x}_i, y_i) \mid i = 1, ..., N\}$ where N is the number of measurements in the data. The covariate can be a scalar x_i (in monotherapy) or a vector $\mathbf{x}_i$ (in combination) corresponding to the drug dosages. The response y_i can be defined as cell growth or inhibition of growth, depending on how the data are collected. In this study, we focus on inhibitory datasets, where a large dosage typically results in growth inhibition. In this case, y_i is defined as the percentage of growth-inhibited cells. Nevertheless, our method can be easily modified to accommodate the opposite setting where the drug response is enhancing growth with respect to the dosages.

Fig. 1 (A) illustrates a typical dose-response matrix for a combination from a screening, where the first row and column contain monotherapy data and the remaining entries contain combination data. The core aim of our method is to infer the synergistic potency and efficacy given such a matrix (or a vector in the case of monotherapy). To accomplish this, we designed **SynBa**, a **Ba**yesian framework for the inference of **Syn**ergistic effects of drug combinations. SynBa is defined by a prior distribution $p(\boldsymbol{\theta})$ for the parameters $\boldsymbol{\theta}$ and a likelihood function $p(Y \mid \boldsymbol{\theta}, X)$ for the drug responses Y. The likelihood function describes the probability of the responses given the dosage data X and the parameters $\boldsymbol{\theta}$. The prior distribution encodes the existing belief or knowledge about the parameters of interest. In the case of monotherapy, we have $\boldsymbol{\theta} = \{E_0, E_1, C, H, \sigma\}$, whereas in combination, we have $\boldsymbol{\theta} = \{E_0, E_1, E_2, E_3, C_1, C_2, H_1, H_2, \alpha, \sigma\}$. The likelihood function and parameters together encode the shape of the dose-response curve (defined in Boxes 1 and 2 and described in more detail below).

In addition, our method provides a way of predicting unseen dose-response data for both monotherapies and combinations. For example in the case of Fig. 1 (B), the quantity of interest would be the posterior predictive distribution of the response of interest, i.e.

$$p(\tilde{y} \mid X, Y, \tilde{\mathbf{x}}) = \int_{\boldsymbol{\theta}} p(\tilde{y} \mid \boldsymbol{\theta}, \tilde{\mathbf{x}}) p(\boldsymbol{\theta} \mid X, Y) d\boldsymbol{\theta} \tag{1}$$

where $\tilde{\mathbf{x}}$ is the untested dosage of interest, $\tilde{y}$ is the predicted response and $p(\boldsymbol{\theta} \mid X, Y)$ is the posterior distribution over the parameters $\boldsymbol{\theta}$ given training data (X, Y).

Overview of SynBa: Monotherapies

We begin by defining SynBa for monotherapy screens, where the dosage x is a scalar. The likelihood function for the response y is based on the Hill equation [15], which has been the classic choice to model pharmacology data:

$$E(x) = E(0) + \frac{E(\infty) - E(0)}{1 + (\frac{C}{x})^H} \tag{2}$$

where x is the dosage of the drug, $E(x)$ is the corresponding measured response, and H controls the slope of the curve. The interpretation of C depends on the problem and the dataset of interest. In this study, as the focus is on inhibitory datasets, C represents the half maximal inhibitory concentration, which is the dosage required to inhibit the given biological process or biological component by 50%, known as IC_{50}. C quantifies the potency of the monotherapy in the study.

The base level of the monotherapy is denoted by $E_0 := E(0)$, which is the response when no drug is applied. The efficacy of the monotherapy is quantified by $E_1 := E(\infty)$, or denoted as $\mathrm{E}_{\mathrm{inf}}$, which is the maximal inhibition when a sufficiently large dosage x is applied.

Box 1 defines the prior distribution for the parameters and the likelihood model for the response given the dosages, which are also illustrated in Fig. 2.

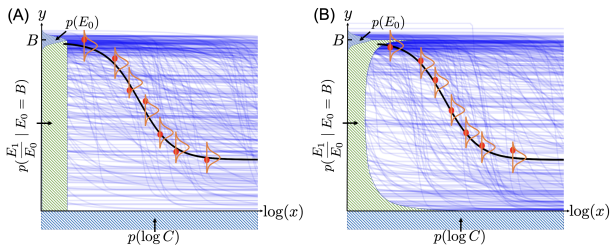

Fig. 2. Two options (A) and (B) for the monotherapy prior model. The prior for $\log C$ is plotted underneath the x-axis, which is uniform in both options. The priors for E_0 (which is Gaussian in both options) and the normalised E_1 are plotted along the y-axis. Given E_0, the normalised E_1 is uniform in Option (A) and follows a Beta(0.46, 0.58) distribution in Option (B). The 300 blue curves are random samples from the expected prior responses $\mathbb{E}[Y|\boldsymbol{\theta}]$ where $\boldsymbol{\theta}$ are sampled from the prior distributions defined in Box 1. The seven red points illustrate an example set of monotherapy dose-response data $\mathcal{D}$. The black curve is a sample from the expected posterior responses $\mathbb{E}[Y|\boldsymbol{\theta}, \mathcal{D}]$ after the model is fitted to the data $\mathcal{D}$, whilst the orange bell-shaped curves illustrate the i.i.d. Gaussian noise for the responses.

To account for observational noise in the data, we define a noise model for y centred around $E(x)$. More specifically, for each fixed experiment setting (i.e. a fixed cell line treated by a fixed monotherapy or combination in the same laboratory environment), y is modelled as having Gaussian noise that is conditionally independent given the dosage: $y \sim \mathcal{N}(E(x), \sigma^2)$. The assumption of conditional independence is valid because in screenings such as the AstraZeneca-Sanger DREAM Challenge [1], the measurements for each monotherapy or combination are performed independently in different plates, instead of being performed sequentially. Therefore, given a set of measurements for a monotherapy or a combination, the noise level of their corresponding responses is independent. The i.i.d. noise is illustrated as orange bell-shaped curves in Fig. 2. As a result, the likelihood function for y is defined as Eq. (4).

The datasets are normalised by measuring cell inhibition when no drug is added (at dose zero). However, due to the noise in the biological process, the inhibition at dose zero would not be the same if the experiment is repeated multiple times. Thus, the normalization procedure itself contains uncertainty. Therefore, we define E_0 to be probabilistic instead of a fixed initial value. A Gaussian prior with mean B and variance $0.03B$ is given for E_0, where B is the normalized inhibition of a dose zero, e.g. $B = 100$ in the DREAM dataset. The variance of this prior is defined to be $0.03B$ so that it is flexible enough to allow for errors, but not too conservative at the same time.

The prior for the normalised maximal response (i.e. $\tilde{E}_1 := \frac{E_1}{E_0}$, with a range of $[0, 1]$) may be defined in various ways depending on whether to insert existing knowledge or historical information. One option is to remain uninformative and impose a uniform prior, as shown in Fig. 2 (A).

Alternatively, we may make use of the existing information from the monotherapy data available. According to the single agent datasets in DREAM, the empirical distribution of $\mathrm{E}_{\mathrm{inf}}$ has a high density on both extremes of the range, with 17.9% of them smaller than 0.05 and 15.1% equal to 1 (after normalising to the interval $[0, 1]$). Using maximum likelihood estimation to fit a $\mathrm{Beta}(a, b)$ distribution to these $\mathrm{E}_{\mathrm{inf}}$ values, we obtain $a = 0.46$ and $b = 0.58$. To account for this information, we define $\mathrm{Beta}(0.46, 0.58)$ as the second option of the prior for the normalised maximal response, as illustrated in Fig. 2 (B). This prior is consistent with the biological behaviour that a drug administered with a sufficiently large dose will either kill the majority of the targeted cells if effective, or very few of them if not effective. The choice of this empirical prior shows how existing knowledge or information on the dynamic/kinetic of drugs can be conveniently added to SynBa through its priors.

A uniform prior is imposed for the logarithm of IC_{50}, or $\log C$, with C bounded by δ and M. The values of δ and M depend on the dataset and the unit of the dosages. In this work, we define δ to be smaller than any non-zero dosage in the dataset we model and define M to be larger than any non-zero dosage in the dataset. In another word, we have $0 < \delta < \min\{x \mid x > 0, (x, y) \in \mathcal{D}\}$ and $M > \max\{x \mid (x, y) \in \mathcal{D}\}$ For example, $\delta = 10^{-10}$ and $M = 10^6$ is a viable choice for DREAM, whereas in NCI-ALMANAC, we may have $\delta = 10^{-15}$ and $M = 10$. The idea is to be uninformative about $\log C$ a priori, due to the fact that the ideal dosage range for the experiments is unknown and often unsuitable, either too small or too large. It is common that IC_{50} exceeds the maximum dosage, which is challenging to infer from the data. In other cases, IC_{50} may lie between zero dosage and the smallest non-zero dosage, due to the tested dosages being too large. Our method includes these possibilities a priori.

H and σ are both given a $\mathrm{lognormal}(0, 1)$ prior because they are both non-negative and assumed to be moderately small. In addition, previous literature has agreed that H approximately follows a lognormal distribution [9]. A $\mathrm{lognormal}(0, 1)$ prior ensures that $P(H < 5) \approx 0.95$ and $P(\sigma < 5) \approx 0.95$ a priori.

The blue curves in Fig. 2 are 300 random samples from the expected prior responses $\mathbb{E}[Y|\boldsymbol{\theta}]$ where $\boldsymbol{\theta}$ are sampled from the prior distributions. It can be observed that the curves cover a wide range of possibilities a priori. Yet, they are not excessively general, as all of them follow the Hill equation. For comparison, we also show (in black) an example of a sample from the posterior distribution given a dataset of seven measurements, and we see how the posterior sample falls close to the data.

Note that the specific priors we have chosen have been motivated in a generic fashion through general knowledge and previous literature. When more knowledge exists for a specific combination, the prior can be conveniently adjusted

Box 1: Overview of SynBa for the inference of monotherapy dose-response data.

The joint prior distribution for the response and the parameters of the curve given the dosages is

$$p(Y, E_0, E_1, C, H, \sigma \mid x) = p(Y \mid E_0, E_1, C, H, \sigma, x)p(E_1 \mid E_0)p(E_0)p(C)p(H)p(\sigma). \tag{3}$$

where the likelihood model for the responses Y given the dosage x is defined as

$$p(y \mid E_0, E_1, C, H, \sigma, x) = \mathcal{N}(E_0 + \frac{E_1 - E_0}{1 + (\frac{C}{x})^H}, \sigma^2) \tag{4}$$

independently for all $y \in Y$, where $E_0 := E(0)$, $E_1 := E(\infty)$ and σ is the standard deviation of the noise level of y, and the priors are

$$\begin{aligned} &p(E_0) = \mathcal{N}(B, 0.03B), \qquad p(\frac{E_1}{E_0}) = \mathcal{U}(0,1) \ \underline{\text{or}} \ \text{Beta}(0.46, 0.58), \\ &p(\log(C)) = \mathcal{U}(\log(\delta), \log(M)), \qquad p(H) = \text{lognormal}(0,1), \qquad p(\sigma) = \text{lognormal}(0,1), \end{aligned} \tag{5}$$

where $0 = x_1 < x_2 \leq x_3 \leq ... \leq x_n$ are the dosages, and $\delta \in (0, x_2)$ is a small non-zero value to avoid $\log C$ being undefined.

to accommodate this, since the inference framework is agnostic to the choice of prior.

Fig. 3 is an illustrative example of SynBa trained on a monotherapy (the compound MTOR_1 treated on the cell line MDA-MB-231) with six measurements, taken from the DREAM dataset. The first row shows the resulting model with a uniform prior for the normalised $\mathrm{E_{inf}}$ (i.e. $\tilde{E}_1$), whereas the second row shows the model with the Beta$(0.46, 0.58)$ prior for $\tilde{E}_1$. We start with two measurements and add two additional responses each time. It can be observed that the posterior distribution for $\mathrm{IC_{50}}$ narrows down quickly for both models because the observed responses span across the range between 40 and 100, which provides sufficient information to estimate $\mathrm{IC_{50}}$ with low uncertainty. The posterior for $\mathrm{E_{inf}}$, on the other hand, is more uncertain, due to the dosage range being too small to observe the convergence of the responses. In this case, the two models provide a visibly different posterior distribution for $\mathrm{E_{inf}}$, due to the different priors. As shown in Fig. 3 (F), the maximum a posteriori probability estimate for $\mathrm{E_{inf}}$ is 0 when the Beta prior is imposed, which results from the inserted prior knowledge that the response is likely to converge to 0 if an effective (but non-zero) response is already observed. This example shows how the prior design affects the inference of the parameter uncertainty. Nevertheless, if we look at the posterior predictive distribution for the responses (with samples illustrated by light blue curves), the two models reach a similar conclusion. We will show in Result that the predictive performance of the model is insensitive to the different choices of prior.

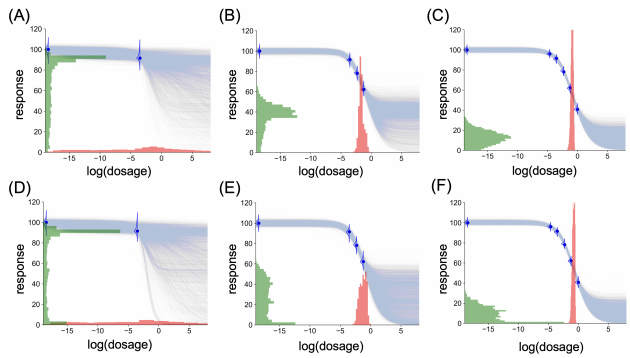

Fig. 3. An example illustration of SynBa on a set of monotherapy data with six measurements. Each blue curve is a sample from $\mathbb{E}[Y \mid \boldsymbol{\theta}]$ where $\boldsymbol{\theta} \sim p(\boldsymbol{\theta} \mid \mathcal{D})$. The distribution for $\mathrm{IC_{50}}$ is shown in red, whereas the distribution for $\mathrm{E_{inf}}$ is shown in green. (A)-(C): The prior distribution for $\tilde{E}_1$ is $\mathcal{U}(0,1)$. (D)-(F): The prior distribution for $\tilde{E}_1$ is Beta$(0.46, 0.58)$.

Overview of SynBa: Combinations

Extending SynBa to combinations of two drugs, the goal is now to model the dose-response surface $E(x_1, x_2) = f(x_1, x_2)$ where x_1 and x_2 are the dosages for the two drugs respectively, whereas $E(x_1, x_2)$ is the response, and f is some class of function to be defined.

To define our likelihood model for the responses Y, we take inspiration from [8, 9] but maintain some flexibility on their model assumptions. The effect of a drug in a system is usually described by the Hill equation that describes the state of equilibrium of a reversible process between an unaffected population and an affected one (the principle of *detailed balance*). To obey to this equation and its effects, in our model we incorporate the assumptions of the principle of detailed balance, of the proliferation rate of unaffected population and of the saturation of the maximum effect of the drug in the affected population. In MuSyC these assumptions are defined in a nested structure, which levels are called *Tiers* (see Table S5 in [8]). In defining SynBa for combinations we are adopting the same model assumptions as MuSyC, using the category *Tier 4* of the levels specified by MuSyC models. This category encodes the most complex class of models that still maintains the assumption of *detailed balance*. It is worth noting that, conversely to MuSyC, our model posterior covers all four tiers simultaneously. This is because *Tier 4* subsumes all lower tiers and thus for how our model is defined, they will not be eliminated from the posterior distribution, unless the evidence from the data is strongly against them. The concepts of Tiers as described in MuSyC would require a post-learning model selection. The use of a Bayesian approach avoids any post-learning model selection whilst still maintaining biologically and physically viable assumptions.

We choose to maintain the detailed balance assumption to also avoid over-parameterisation, which is likely to occur due to the limited data size for each drug combination set. For example, the majority of the combinations in the NCI-ALMANAC dataset [2], one of the most widely used in pharmacology, has a data size of 15 (excluding the base level at dose zero), only 3 more than the number of parameters in MuSyC. In a real-world scenario, the data size may often be even smaller. Furthermore, with the detailed balance assumption, matrix multiplication and inversion are avoided, which lowers the computational cost.

Box 2 defines the prior distribution for the parameters and the likelihood model for the response given the dosages.

Box 2: Overview of SynBa for the inference of combination dose-response data.

We define the following joint distribution for the response and the parameters of the surface given the dosages $\mathbf{x} = (x_1, x_2)$,

$$\begin{aligned} &p(Y, E_0, E_1, E_2, E_3, C_1, C_2, H_1, H_2, \sigma, \alpha \mid \mathbf{x}) \\ &= p(Y \mid E_0, E_1, E_2, E_3, C_1, C_2, H_1, H_2, \sigma, \alpha, \mathbf{x}) p(E_1 \mid E_0) p(E_2 \mid E_0) p(E_3 \mid E_0) p(E_0) p(C_1) p(C_2) p(H_1) p(H_2) p(\sigma) p(\alpha), \end{aligned} \tag{6}$$

where the likelihood model for responses Y given the dosages is defined as

$$p(y \mid E_0, E_1, E_2, E_3, C_1, C_2, H_1, H_2, \sigma, \alpha, \mathbf{x}) = \mathcal{N}\left(\frac{C_1^{H_1} C_2^{H_2} E_0 + x_1^{H_1} C_2^{H_2} E_1 + C_1^{H_1} x_2^{H_2} E_2 + \alpha x_1^{H_1} x_2^{H_2} E_3}{C_1^{H_1} C_2^{H_2} + x_1^{H_1} C_2^{H_2} + C_1^{H_1} x_2^{H_2} + \alpha x_1^{H_1} x_2^{H_2}}, \sigma^2\right) \tag{7}$$

for all $y \in Y$, where $E_0 := E(0,0)$, $E_1 := E(\infty, 0)$, $E_2 := E(0, \infty)$, $E_3 := E(\infty, \infty)$, C_1 and H_1 are the monotherapeutic parameters associated for Drug 1, C_2 and H_2 are the monotherapeutic parameters associated for Drug 2, and α is an association parameter that controls how the two drugs are affected by the presence of each other.

The priors are

$$\begin{aligned} &p(E_0) = \mathcal{N}(B, 0.03B), \qquad p\left(\frac{E_i}{E_0}\right) = \mathcal{U}(0,1) \ \underline{\text{or}}\ \text{Beta}(0.46, 0.58) \quad \text{for } i = 1,2,3, \\ &p(\log(C_i)) = \mathcal{U}(\log(\delta), \log(M)) \ \text{ and } \ p(H_i) = \text{lognormal}(0,1) \quad \text{for } i = 1,2, \text{ and} \\ &p(\alpha) = \text{lognormal}(0,1), \qquad p(\sigma) = \text{lognormal}(0,1), \end{aligned} \tag{8}$$

where $0 = x_{i,1} \leq x_{i,2} \leq \ldots \leq x_{i,n}$ are the dosages for drug i, and δ is a small non-zero value to avoid $\log C$ being undefined.

The definitions of the priors are a natural extension from the monotherapy model, with the same arguments being followed. The only new parameter is α, which follows a lognormal prior with median 1 because α is non-negative and equals 1 when the combination is additive (i.e. neither synergistic nor antagonistic) in terms of potency.

Inference of the synergy

After inferring the posterior for the parameters and their associated uncertainty, we focus on distinguishing the effect of efficacy and potency in drug combinations. MuSyC has defined metrics for both synergistic efficacy and synergistic potency, which is a promising step in decoupling potency and efficacy. However, the uncertainty for these two quantities has not been quantified systematically. Our model output includes not only quantification for the synergistic efficacy and the synergistic potency, but also a separate uncertainty estimation for each of them.

For the synergistic efficacy, one simple yet informative quantity is $\Delta\text{HSA} = \min(E_1, E_2) - E_3$ which is the change in the maximal effect between the combination and the more effective single drug of the two [16]. A positive score indicates synergistic efficacy. As E_1, E_2 and E_3 are probabilistic, the resulting ΔHSA score is also probabilistic. A metric such as

$$P(\Delta\text{HSA} > 0 \mid \mathcal{D}) \tag{9}$$

can then be defined to estimate how confident we are about the *synergistic efficacy* of the combination, based on the dataset $\mathcal{D}$. It is possible to have a synergistic combination that is highly uncertain, which would indicate that more data are required to reach confidence in the estimation.

For the synergistic potency, α contains the required information. $\alpha > 1$ would indicate synergistic potency [9], which means the potency of the two drugs has reduced due to being combined. Consequently, we define

$$P(\alpha > 1 \mid \mathcal{D}) \tag{10}$$

as an estimation of how likely the combination satisfies *synergistic potency*.

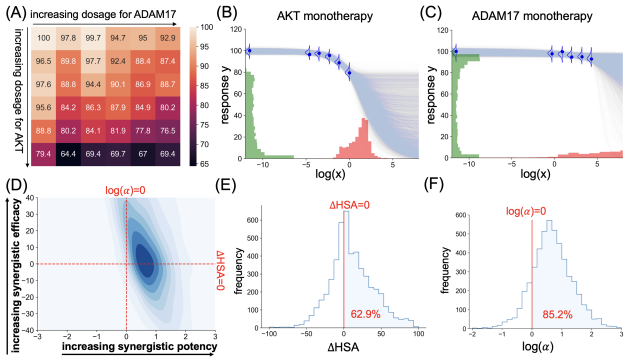

Fig. 4. The data and the inference outputs of the combination of AKT and ADAM17 applied on the cell line BT-20. (A): The original dose-response matrix. (B)-(C): The monotherapy model outputs. Each blue curve is a sample from $\mathbb{E}[Y \mid \boldsymbol{\theta}]$ where $\boldsymbol{\theta}$ is a sample from the posterior of the respective monotherapy model. The posterior distribution for IC_{50} and E_{inf} are shown in red and green respectively. (D): The contour plot for the joint posterior distribution of the synergistic efficacy (ΔHSA) and the synergistic potency ($\log(\alpha)$). The distribution is smoothed from the empirical posterior with a kernel density estimation for visualisation purpose. (E)-(F): The histogram of the empirical posterior distribution for the synergistic efficacy (ΔHSA) and the synergistic potency ($\log(\alpha)$) respectively. The areas on the right hand side of the red vertical lines are the probability that the combination is synergistic in terms of efficacy (in (E)) and potency (in (F)).

Case studies

To illustrate how the uncertainty estimation from our method can be explained and further used for decision-making, we take two combinations from the DREAM dataset [1] as examples.

Fig. 4 (A) shows the dose-response matrix for the combination of ADAM17 and AKT acted on the cell line BT-20. The first column is the monotherapy dose-response data for AKT (as the dosage for ADAM17 is zero), whilst the first row is the monotherapy dose-response data for ADAM17 (as the dosage for AKT is zero). It can be observed that the responses for AKT start to decrease at a higher rate when the dosage increases, but the dosage range is too small to

understand its potency (IC_{50}) and efficacy (E_{inf}). The efficacy cannot be determined because the response has not shown any sign of convergence, whereas the potency cannot be determined because it relies on understanding the maximal response, which is itself uncertain. However, if a deterministic Hill equation is fitted to the monotherapy data, it will only provide a point estimate for the IC_{50} and E_{inf}, without acknowledging the above caveats. On the contrary, our method provides an uncertainty estimation for both quantities. As shown in Fig. 4 (B)-(C), the posterior distribution of IC_{50} and E_{inf} have large variances, which correspond to large uncertainty. In particular, as shown in Fig. 4 (C), the posterior of E_{inf} for ADAM17 has a multimodal shape, which is sensible because it is unclear whether the dosage range is too small (which corresponds to the peak at 0), or the drug is simply ineffective regardless of the dosage (which corresponds to the peak at around 90).

Moving to the inference of the full combination matrix, most existing synergy methods have no means to showcase the uncertainty. Our method, on the contrary, provides the uncertainty around the synergistic potency and the synergistic efficacy, as shown in Fig. 4 (D), (E) and (F). According to the model output, the combination is moderately likely to be synergistically potent (with a probability of 85.2%), but it is difficult to conclude its synergistic efficacy (with a probability of 62.9% to be synergistic effective). This is reasonable because the excessively small dosage range makes it impossible to conclude anything about efficacy with low uncertainty, but with the 25 available measurements on the plates where the two drugs have interacted, information can be extracted on whether combining the two drugs may lower the level of toxicity required to reach the same beneficial effect.

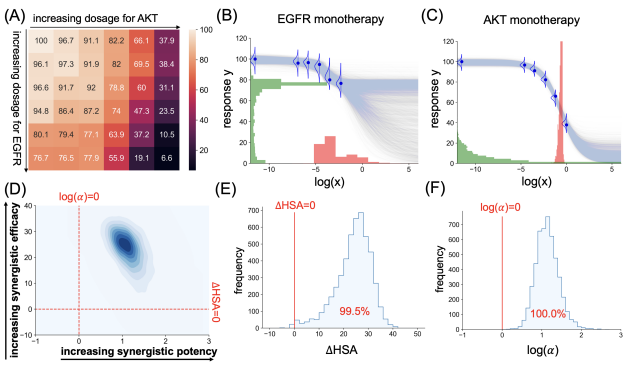

Fig. 5. The data and the inference outputs of the combination of EGFR and AKT applied on the cell line MDA-MB-468. (A)-(F): The same as the caption of Fig. 4.

This combination is an example where the model implies some potential in the synergy of the combination, but the level of uncertainty in the synergy is still high, which may require more measurements at larger dosages to be narrowed down. The uncertainty captured in SynBa allows conclusions such as this to be drawn, which is a helpful guidance for subsequent experiments.

We now consider the combination of AKT and EGFR acted on the cell line MDA-MB-468. Fig. 5 shows its dose-response matrix and the inference result from our model for the monotherapies and the combination respectively. All parameters and metrics of interest have low variances, representing low uncertainties. As shown in Fig. 5 (B) and (C), the dosages have suitable ranges and approximately follow the sigmoidal shapes of the expected dose-response fit, in particular for AKT. They contain sufficient information for the possibilities for IC_{50} and E_{inf} to be narrowed down. Similarly, the combination data are well-behaved. Fig. 5 (C), (D) and (E) show that the probabilities of this combination being synergistic in terms of potency and efficacy are both close to 100%. These are signs that this combination is worth being taken to subsequent steps in the drug development pipeline.

The two examples above show that concrete decisions can be made based on the posterior distributions (e.g. for IC_{50}, E_{inf}, ΔHSA and α) from our model, and more importantly, the uncertainties associated with these distributions.

Training details

The models are trained by Stan [17], a state-of-the-art platform for statistical modelling and high-performance statistical computation, particularly for Bayesian computation. The user specifies the prior model of the parameters and the likelihood model of the data, while Stan performs either full Bayesian statistical inference with Markov chain Monte Carlo (MCMC) sampling, or approximate Bayesian inference with variational inference. While full Bayesian inference with MCMC sampling provides an asymptotically exact solution, it is slower and has worse scalability than variational inference [18, 19]. In this study we use MCMC due to the small size of the data and the model, as well as the importance of the reliability of the output, which is strengthened by the asymptotic exactness of the MCMC inference. Despite choosing the slower option, SynBa is still computationally efficient. Running on 4 CPUs of the Intel Xeon Platinum 8276 CPU Processor, the median time taken to fit SynBa (via MCMC with 1000 iterations and 4 chains, including 500 iterations in the warm-up phase) to a 6-by-6 dose-response matrix in DREAM is 10.2 seconds, which is comparable to MuSyC with bootstrap.

With this training pipeline, we can avoid the overhead that occurs during the usage of non-linear optimisation packages in deterministic parametric methods such as MuSyC, BRAID and the Effective Dose model. A different choice of the numerical algorithm (and its hyperparameters) results in a different result for those methods. On the contrary, in SynBa, the same exact result can be found asymptotically via MCMC with Stan.

For the implementation of the other benchmark methods including MuSyC, BRAID and the Effective Dose model, the Python package `synergy` [20] is used.

Results

Prediction of drug combination responses

In this subsection, we show that in addition to providing uncertainty estimations, SynBa is competitive in predicting unseen responses within a dose-response matrix, and is less prone to overfitting compared to the existing methods.

The datasets of interest are DREAM [1] and NCI-ALMANAC [2], two of the most widely-used publicly-available combination screenings. In DREAM, we focus on all examples in the training set of Challenge 1 that have passed the Quality Assurance and that only contain non-negative responses and one set of replicates, which are 1631 sets of combinations in total. In NCI-ALMANAC, we focus on the subset defined in [21], which is a subset of the data consisting of 50 unique FDA-approved drugs and 36,120 combinations in total. We remove

examples that contain negative measurements, which results in 28,854 remaining combinations to be experimented on.

We leave out 20% of the non-zero dosage combinations for prediction. The models are trained using the remaining 80% dosage combinations, and then evaluated on the left-out points. For the dose-response matrices in DREAM, we leave out 7 of the 35 points with non-zero dosages from the 6-by-6 matrix (see Fig. 1(A)) using a specific leave-out strategy. For each of the two monotherapy slices, one point is left out for testing. For the 5-by-5 combination grid (i.e. the orange cells in Fig. 1(A)), five points are randomly left out for testing. Fig. 1(B) shows an example of what a train-test split may look like for DREAM. The measurements are left out in this manner so that each monotherapy contains one measurement for testing.

For the dose-response matrices in NCI-ALMANAC, it is not possible to leave out points separately for monotherapies and for combinations because the data size is too small. For most combinations, there are only 3 points for each monotherapy and 9 points for the interactions. Thus, we directly leave out 3 of the 15 points randomly for prediction.

Evaluation metrics

To evaluate the predictive performance of the models, test likelihood and the root-mean-square error (RMSE) are computed using the left-out points. Both are common metrics for evaluating the goodness of fit. The former focuses on the goodness of the full predictive distribution, whereas the latter focuses solely on the goodness of the point estimates for the responses.

The computation of the test likelihood for SynBa follows Equation (1) for both monotherapy and combination. However, the right hand side of the equation cannot be computed in a closed form, so Monte Carlo estimation is required using the expression

$$p(\tilde{y} \mid \mathcal{D}, \tilde{\mathbf{x}}) \approx \frac{1}{M} \sum_{m=1}^{M} p(\tilde{y} \mid \boldsymbol{\theta}_m, \tilde{\mathbf{x}}) \tag{11}$$

where $\tilde{y}$ is the predicted response of the left-out dosage, $\mathcal{D}$ is the training data containing the known dosages and the known responses, and $\boldsymbol{\theta}_m \sim p(\boldsymbol{\theta} \mid \mathcal{D})$ are the MCMC samples from the posterior distribution for the parameters.

For MuSyC, BRAID and the Effective Dose model, the parametric bootstrap pipeline described in [9] is followed. Each bootstrapped dataset provides a fitted curve. The density for $\tilde{y}$ is then estimated by averaging its density computed on the models learnt from the bootstrapped datasets.

The computation of RMSE is more straightforward. For each combination, its RMSE for the test responses $\{y_1, ..., y_N\}$ is

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{12}$$

where $\hat{y}_i$ is the point estimate for the response that corresponds to dosage $\mathbf{x}_i$. For SynBa, we define $\hat{y}_i$ to be the posterior predictive mean $\mathbb{E}[y_i \mid \mathcal{D}, \mathbf{x}_i]$, which is estimated by the Monte Carlo estimator

$$\mathbb{E}[y_i \mid \mathcal{D}, \mathbf{x}_i] \approx \frac{1}{M} \sum_{m=1}^{M} \mathbb{E}[y_i \mid \boldsymbol{\theta}_m, \mathbf{x}_i] \tag{13}$$

where $\boldsymbol{\theta}_m \sim p(\boldsymbol{\theta} \mid \mathcal{D})$ are the MCMC samples from the posterior, and $\mathbb{E}[y_i \mid \boldsymbol{\theta}_m, \mathbf{x}_i]$ can be computed by Eq. (7).

Quantitative results

We compare our prediction results against MuSyC, BRAID, and the Effective Dose model, which are three of the most widely-used synergy models. For SynBa, we implement both the uniform prior and the empirical Beta prior for the normalised $\mathrm{E}_{\mathrm{inf}}$, which we denote as SynBa-U and SynBa-B respectively. Tables 1 show the mean and the median of the test log-likelihood and the test RMSE for MuSyC, BRAID, Effective Dose model and SynBa. It can be observed that our method outperforms all three other methods in all metrics except for the mean test log-likelihood on DREAM. What is worth noting is that our method performs strongly on RMSE, which is a metric that only considers the quality of point estimates and ignores uncertainty. The upper diagonal panels in Fig. 6 show the scatter plots directly comparing the test RMSE values between methods (visualised with the blue colour). Our method is the most competitive, as evidenced by having more points above the diagonal $y = x$ line. These show that our method is not trading off predictive accuracy for uncertainty estimation. By following a principled Bayesian workflow, our model is strong in both prediction and uncertainty estimation.

It is worth noting that at least one of MuSyC, BRAID or the Effective Dose model fail to find a solution for 4.8% of the examples in DREAM and 38.4% of them in NCI-ALMANAC, despite an effort in tuning the bounds, initial values and hyperparameters involved in the optimisation. Most likely this is because these methods rely on external optimisation packages with no guaranteed convergence, which can become a problem when overparameterisation becomes severe due to small data sizes. SynBa does not incur this problem since its priors ensure conservative outputs when data size is too small.

To investigate whether SynBa is prone to overfitting and how it compares to the other three methods, we perform the same prediction experiment on DREAM, but with a train-test split ratio of 40% : 60% instead, so that we may observe how the test RMSE changes on exactly the same data but with a different train-test split ratio.

As shown in the lower diagonal panels in Fig. 6, the test RMSE values (visualised with the red colour) increase significantly for MuSyC, BRAID and the Effective Dose model. For SynBa, however, the test RMSE values have increased on average, but not by much. It can be seen that the mean value and the spread increase more significantly for the other three methods compared to SynBa. It can also be observed that the predictive performance of SynBa is not sensitive to the choice of prior, with the two priors producing very close RMSE values to each other.

Uncertainty calibration

For a model $\mathcal{M}$ with learnt cumulative distribution $F_{\mathcal{M}}$ with well-calibrated uncertainty, it would approximately follow the identity that

$$F_Y(\mathbf{x}_i) \approx F_{\mathcal{M}}(\mathbf{x}_i) \tag{14}$$

for every data point $\{(\mathbf{x}_i, y_i(\mathbf{x}_i)) \mid i = 1, ..., N\}$ in the dataset, where $y_i(\mathbf{x}_i)$ is a sample from the unknown true cumulative distribution $F_Y(\mathbf{x}_i)$. Equivalently, assuming the measurements $y_i(\mathbf{x}_i)$ for a combination are conditionally independent given the dosages $\mathbf{x}_i$, their cumulative probabilities $F(y_i) := \mathbb{P}(y_i(\mathbf{x}_i) < F_{\mathcal{M}}(\mathbf{x}_i))$ would be approximately uniformly distributed between 0 and 1, if $\mathcal{M}$ is well-calibrated.

In this study, for each combination in DREAM, we split the 35 measurements (excluding the base value) with a 80%:20% ratio in the same way as the prediction evaluation in the previous subsection. We then evaluate the quality of the uncertainty calibration with the Kolmogorov–Smirnov (K-S) uniformity test [22] for the empirical cumulative probabilities

Table 1. The mean and the median (Mdn) of the test log-likelihood (LL) and the test root-mean-squared error (RMSE) for MuSyC, BRAID, the Effective Dose model (ED) and SynBa, computed on a subset of DREAM and NCI-ALMANAC, along with their standard errors. The standard error of the mean is computed by the standard deviation of the metrics across examples divided by the square root of the number of examples. The standard error of the median is estimated by nonparametric bootstrap. SynBa with a uniform prior for the normalised E_{inf} is denoted by SynBa-U. SynBa with the Beta(0.46, 0.58) prior for the normalised E_{inf} is denoted by SynBa-B.

	DREAM, LL		DREAM, RMSE		NCI-ALMANAC, LL		NCI-ALMANAC, RMSE	
	mean (± se)	Mdn (± se)	mean (± se)	Mdn (± se)	mean (± se)	Mdn (± se)	mean (± se)	Mdn (± se)
MuSyC	-3.50 ± 0.35	-3.09 ± 0.02	6.11 ± 0.10	5.01 ± 0.08	-3.86 ± 0.01	-3.79 ± 0.01	14.87 ± 0.10	10.17 ± 0.03
BRAID	-4.12 ± 0.69	-3.06 ± 0.02	5.71 ± 0.09	4.88 ± 0.08	-3.80 ± 0.04	-3.48 ± 0.01	9.57 ± 0.06	7.16 ± 0.03
ED	**-3.37** ± 0.09	-3.22 ± 0.02	6.46 ± 0.09	5.66 ± 0.09	-3.68 ± 0.02	-3.48 ± 0.01	8.47 ± 0.05	6.77 ± 0.09
SynBa-U	-3.59 ± 0.42	**-3.01** ± 0.02	5.20 ± 0.07	4.56 ± 0.08	-3.42 ± 0.02	-3.24 ± 0.01	6.72 ± 0.04	5.45 ± 0.05
SynBa-B	-3.84 ± 0.66	**-3.01** ± 0.02	**5.15** ± 0.07	**4.55** ± 0.08	**-3.39** ± 0.01	**-3.23** ± 0.01	**6.66** ± 0.04	**5.43** ± 0.04

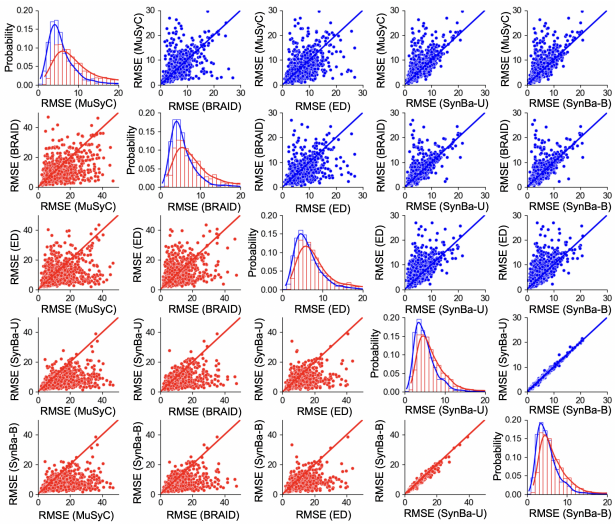

Fig. 6. Upper diagonal panels (with blue points): Scatter plot of the test RMSE values obtained from different methods on DREAM with a train-test split ratio of 80% : 20%. Lower diagonal panels (with red points): Scatter plot of the test RMSE values obtained from different methods on DREAM with a train-test split ratio of 40% : 60%. Diagonal panels: Histograms of the test RMSE values and their corresponding kernel density estimates, where the blue ones represent a train-test split ratio of 4:1, whilst the red ones represent a train-test split ratio of 2:3.

(or CDF values) across all test data points. If the model is well-calibrated, then the CDF values for the test data points will be approximately uniformly distributed for each combination. Otherwise, they will show a non-uniform pattern, and the resulting p-value for the K-S test will be statistically significant. This procedure is performed across every combination in DREAM, resulting in a p-value for each combination. The resulting histogram of the p-values for SynBa is in Fig. 7 (A), showing that 6.07% of the combinations have not passed the uniformity test, and thus are not well-calibrated. As a comparison, the same procedure is performed over DREAM using MuSyC. Fig. 7 (B) shows that 25.1% of the combinations are not well-calibrated when MuSyC is chosen as the model, which is roughly four times as high as the number for SynBa. This shows that SynBa provides a better-calibrated uncertainty, which means its estimated uncertainty is more reliable and closer to the unknown ground truth on average.

Discussion

Machine learning methods have been developed for preclinical modelling and the prediction of drug combinations, thanks

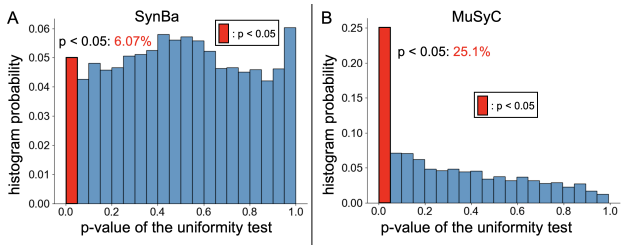

Fig. 7. Histogram of p-values that represent the calibration quality of combination models. The p-values are derived from the Kolmogorov–Smirnov test between the uniform distribution and the cumulative probability of the data points in their predicted densities. For a well-calibrated model, the p-value will be higher than 0.05.

to the availability of large screenings in recent years [21, 23]. The development of such methods is beneficial for discovering and explaining drug combinations. However, a few factors prevent most of these methods from being applied to real-world drug discovery projects [24]. First, most of them are trained to perform well on datasets with a set format and do not generalise well to real-world scenarios where the need of defining decision criteria for choosing optimal drug combinations is prominent. Second, the performance measure is based on synergy scores, which is a proxy measure that contains its shortcomings, as discussed above. Due to the limited data size for each combination and the inherent noise in the data, any performance measure would be associated with a non-trivial amount of uncertainty. However, uncertainty measurement is not included in the estimation of synergy scores. The Spearman correlation of the replicate experiments performed in DREAM [1] and the O'Neil et al. dataset [25] are 0.56 and 0.63 respectively, which show that quantifying a combination with a single synergy score would result in a high variance. A high level of uncertainty in the data is ignored during the modelling process. This could be one of the reasons that 20% of drug combinations are poorly predicted by all methods in the DREAM challenge [1]. Measuring the uncertainties associated to the estimated scores is important in real-world scenarios for the subsequent decision-making process based on the model outputs. In real-world scenarios, scientists are often facing the decision to choose amongst a large set of drug combinations that score similarly in terms of synergy. Without any quantification of how certain (or uncertain) the estimated scores are, they will have to rely on background knowledge compromising innovation in their choices. SynBa provides a way to implement a ranking strategy in the decision process of a drug-discovery pipeline, which is a real-world unmet need.

Conclusions

We have developed a new framework for quantifying dose-response relationships for monotherapies and combinations that provides a full uncertainty estimation for all parameters that are associated with the monotherapies and the combinations, including information about efficacy, potency and synergy. These uncertainty information would be helpful to the biologists to make further decisions about progressing to the next stages of the drug discovery pipeline, or whether more experiments are required to lower the level of the uncertainty and better understand the drug mechanism of action.

We have also shown that SynBa is competitive in predicting unseen responses within a given dose-response matrix, and outperforms MuSyC, BRAID and the Effective Dose model on DREAM and NCI-ALMANAC. In addition, the prediction performance is not sensitive to the choice of the priors.

In summary, our framework is capable of (1) providing a reliable uncertainty estimation for the potency (e.g. IC_{50}) and the efficacy (e.g. E_{inf}) of a monotherapy, or the synergistic potency and efficacy of a combination, in a decoupled manner, and (2) reliably predicting unseen responses within a dose-response matrix. The parameter uncertainties can be interpreted and used as guidance for further experiments and subsequent decision-making.

Code and Data Availability

The method implementation is available at `https://github.com/HaotingZhang1/SynBa`. The datasets used in this study are publicly available (DOI of DREAM: 10.7303/syn4231880; DOI of the subset of NCI-ALMANAC: 10.5281/zenodo.4135059).

Acknowledgments

The authors thank Mateja Jamnik, Pietro Liò, Krishna Bulusu and Paul Metcalfe for their valuable comments and suggestions. This work was performed using resources provided by the Cambridge Service for Data Driven Discovery (CSD3) operated by the University of Cambridge Research Computing Service, provided by Dell EMC and Intel using Tier-2 funding from the Engineering and Physical Sciences Research Council (capital grant EP/T022159/1), and DiRAC funding from the Science and Technology Facilities Council.

Funding

HZ acknowledges the receipt of studentship award from the Health Data Research UK-The Alan Turing Institute Wellcome PhD Programme in Health Data Science (Grant Ref: 218529/Z/19/Z) and the Wellcome Cambridge Trust Scholarship. MR is supported by a Wellcome Trust award (204832/B/16/Z).

References

1. M. P. Menden et al. Community assessment to advance computational prediction of cancer drug combinations in a pharmacogenomic screen. *Nature Communications*, 10(1):1–17, 2019.
2. S. L. Holbeck et al. The national cancer institute ALMANAC: A comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity. *Cancer Research*, 77(13):3564–3576, 2017.
3. B. Zagidullin et al. DrugComb: an integrative cancer drug combination data portal. *Nucleic acids research*, 47(W1):W43–W51, 2019.
4. P. Jaaks et al. Effective drug combinations in breast, colon and pancreatic cancer cells. *Nature*, 603(7899):166–173, 2022.
5. K. C. Bulusu et al. Modelling of compound combination effects and applications to efficacy and toxicity: state-of-the-art, challenges and perspectives. *Drug discovery today*, 21(2):225–238, 2016.
6. C. I. Bliss. The toxicity of poisons applied jointly 1. *Annals of applied biology*, 26(3):585–615, 1939.
7. S. Loewe. The problem of synergism and antagonism of combined drugs. *Arzneimittelforschung*, 3:285–290, 1953.
8. C. T. Meyer et al. Quantifying drug combination synergy along potency and efficacy axes. *Cell systems*, 8(2):97–108, 2019.
9. D. Wooten et al. MuSyC is a consensus framework that unifies multi-drug synergy metrics for combinatorial drug discovery. *Nature communications*, 12(1):1–16, 2021.
10. N. R. Twarog et al. BRAID: a unifying paradigm for the analysis of combined drug action. *Scientific Reports*, 6(1):1–15, 2016.
11. A. Zimmer et al. Prediction of multidimensional drug dose responses based on measurements of drug pairs. *Proc. Natl. Acad. Sci.*, 113(37):10442–10447, 2016.
12. J. Tang et al. What is synergy? the saariselkä agreement revisited. *Frontiers in pharmacology*, 6:181, 2015.
13. Y. Shapovalova et al. Non-parametric synergy modeling with gaussian processes. *bioRxiv*, 2021.
14. L. Rønneberg et al. bayesynergy: flexible bayesian modelling of synergistic interaction effects in in vitro drug combination experiments. *Briefings in bioinformatics*, 22(6):bbab251, 2021.
15. A. V. Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol.*, 40:4–7, 1910.
16. W. R. Greco. The search for synergy: a critical review from a response surface perspective. *Pharmacol. Rev.*, 47:331–385, 1995.
17. Stan Development Team. Stan modeling language users guide and reference manual, version 2.29, 2022.
18. D. M. Blei et al. Variational inference: A review for statisticians. *J. Am. Stat. Assoc.*, 112(518):859–877, 2017.
19. T. Salimans et al. Markov chain monte carlo and variational inference: Bridging the gap. In *International conference on machine learning*, pages 1218–1226. PMLR, 2015.
20. D. J. Wooten and R. Albert. synergy: a python library for calculating, analyzing and visualizing drug combination synergy. *Bioinformatics*, 37(10):1473–1474, 2021.
21. H. Julkunen et al. Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects. *Nature communications*, 11(1):1–11, 2020.
22. F. J. Massey Jr. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.*, 46(253):68–78, 1951.
23. T. Wang et al. Modeling drug combination effects via latent tensor reconstruction. *Bioinformatics*, 37(Supplement_1):i93–i101, 2021.
24. A. Bender and I. Cortés-Ciriano. Artificial intelligence in drug discovery: what is realistic, what are illusions? part 1: ways to make an impact, and why we are not there yet. *Drug discovery today*, 26(2):511–524, 2021.
25. J. O'Neil et al. An unbiased oncology compound screen to identify novel combination strategies. *Molecular cancer therapeutics*, 15(6):1155–1162, 2016.