bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Additive Dose Response Models: Defining Synergy

Simone Lederer[1], Tjeerd M.H. Dijkstra[2,3], and Tom Heskes[1]

[1]Data Science, Institute for Computing and Intelligent Systems, Radboud University, Postbus 9010, 6500 GL Nijmegen, The Netherlands
[2]Max Planck Institute for Developmental Biology, Max-Planck-Ring 1, 72076 Tübingen, Germany
[3]Center for Integrative Neuroscience, University Clinic Tübingen, Otfried-Müller-Str. 25, 72076 Tübingen, Germany

July 31, 2019

Abstract

In synergy studies, one focuses on compound combinations that promise a synergistic or antagonistic effect. With the help of high-throughput techniques, a huge amount of compound combinations can be screened and filtered for suitable candidates for a more detailed analysis. Those promising candidates are chosen based on the deviance between a measured response and an expected non-interactive response. A non-interactive response is based on a principle of no interaction, such as Loewe Additivity [Loewe, 1928] or Bliss Independence [Bliss, 1939]. In Lederer et al. [2018a], an explicit formulation of the hitherto implicitly defined Loewe Additivity has been introduced, the so-called Explicit Mean Equation. In the current study we show that this Explicit Mean Equation outperforms the original implicit formulation of Loewe Additivity and Bliss Independence when measuring synergy in terms of the deviance between measured and expected response. Further, we show that a deviance based computation of synergy outperforms a parametric approach. We show this on two datasets of compound combinations that are categorized into synergistic, non-interactive and antagonistic [Yadav et al., 2015, Cokol et al., 2011].

Keywords: *synergy, Loewe Additivity, Bliss Independence, dose equivalence, Combination Index, General Isobole Equation, Explicit Mean Equation, Hill curve, null reference model, response surface, lack-of-fit*

1 Introduction

When combining a substance with other substances, one is generally interested in interaction effects. Those interaction effects are usually described as synergistic or antagonistic, dependent on whether the interaction is positive, resulting

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in greater effects than expected, or negative, resulting in smaller effects than expected. From data generated with high-throughput techniques, one is confronted with massive compound interaction screens. From those screens, one needs to filter for interesting candidates that exhibit an interaction effect. To quickly scan all interactions, a simple measure is needed. Based on that preprocessing scan, those filtered combination candidates can then be examined in greater detail.

To determine whether a combination of substances exhibits an interaction effect, it is crucial to determine a non-interactive effect. Only when deviance from that so-called null reference is observed, can one speak of an interactive effect [Lederer et al., 2018a]. Over the last century, many principles of non-interaction have been introduced. For an extensive overview, refer to [Greco et al., 1995, Geary, 2012]. Two main principles for non-interactivity have survived the critics: Loewe Additivity [Loewe, 1928] and Bliss Independence [Bliss, 1939]. The popularity of Loewe Additivity is based on its principle of sham combination which assumes no interaction when a compound is combined with itself. Other null reference models do not hold that assumption. An alternative is Bliss Independence, which assumes (statistical) independence between the combined compounds.

Independent of the indecisive opinions about the null reference, there are multiple proposals how synergy can be measured given a null reference model. Some suggest to measure synergy as the difference between an observed isobole and a reference isobole calculated from a null-reference model. An isobole is the set of all dose combinations of the compounds that reach the same fixed effect, such as 50% of the maximal effect [Minto et al., 2000, Chou and Talalay, 1984]. Another way to quantify synergy on the basis of the isobole is to look at the curvature and arc-length of the longest isobole spanned over the measured response [Cokol et al., 2011]. As the deviation from an isobole is measured for a fixed effect or dose ratio, synergy is only measured locally along that fixed effect or dose ratio. In order to not miss any effects, this method has to be applied for as many dose ratios possible.

In this paper we measure synergy as the deviation over the entire response surface. One way to do so is the Combenefit method by measuring synergy in terms of volume between the expected and measured effect [Di Veroli et al., 2016]. We will refer to it as a lack-of-fit method as it quantifies the lack of fit from the measured data to the null reference model. Another way of capturing the global variation is by introducing a synergy parameter α into the mathematical formulation of the response surface. This parameter α is fitted by minimizing the error between the measured effect and the α-dependent response surface. Such statistical definition of synergy allows for statistical testing of significance of the synergy parameter. Fitting a synergy parameter to the data as in the parametric approaches tends to be computationally more complex than computing the difference between the raw data and the null model as in the lack-of-fit approaches.

There is an increase in theoretical approaches to synergy, such as the recently re-discovered Hand model [Hand, 2000, Sinzger et al., 2019], which is a formulation of Loewe Additivity in form of a differential expression, or new ways of defining and measuring synergy, such as the ZIP model [Yadav et al., 2015], SynergyFinder [He and Tang, 2016], MuSyC [Meyer et al., 2019] and the copula model [Lambert and Dawson, 2019]. It would be a large effort to com-

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

pare these recent approaches with ours. An extensive comparison of the models has recently been made in Meyer et al. [2019]. Hence we focus on the two main principles, Loewe Additivity and Bliss Independence.

As the research area of synergy evolved from different disciplines, different terminologies are in common use. Whilst in pharmacology, one refers to the Loewe model, in toxicology, the same principle is called concentration addition. The response can be measured among others in growth rate, survival, or death. It is usually referred to as the measured or phenotype effect or as cell survival. In this study we interchange the terms response and effect.

When measuring a compound combination, one also measures each agent individually. The dose or concentration is typically some biological compound per unit of weight when using animal or plant models or per unit of volume when using a cell-based assay. However, it can also be an agent of a different type for example a dose of radiation as used in modern combination therapies for cancer [Nat, 2018]. This individual response is called mono-therapeutic response [Di Veroli et al., 2016] or single compound effect. We prefer a more statistical terminology and refer to it as conditional response or conditional effect. With record we refer to all measurements taken of one cell line or organism which is exposed to all combinations of two compounds. In other literature, this is referred to as response matrix [Lehar et al., 2007, Yadav et al., 2015].

In Section 2.1, we give a short introduction to the two null response principles, Loewe Additivity and Bliss Independence. We explain in detail several null reference models that build on those principles. We introduce synergy as any effect different from an interaction free model in Section 2.2. There, we also introduce the parametrized and deviance based synergy approaches. In Section 2.3, we introduce two datasets that come with a categorization into synergistic, non-interactive and antagonistic. We evaluate the models and methods in Section 3 together with a detailed comparison of the synergy scores.

2 Materials and Methods

2.1 Theory

Before one can decide whether a compound combination exhibits a synergistic effect, one needs to decide on the expected effect assuming no interaction between the compounds. Such so-called null reference models are constructed from the conditional (mono-therapeutic) dose-response curves of each of the compounds, which we denote by $f_j(x_j)$ for $j \in \{1, 2\}$. Null reference models extend the conditional dose-response curves to a (null-reference) surface spanned between the two conditional responses. We denote the surface as $f(x_1, x_2)$ such that

$$f(x_1, 0) = f_1(x_1) \tag{1}$$

and

$$f(0, x_2) = f_2(x_2). \tag{2}$$

Thus, the conditional response curves are the boundary conditions of the null reference surface. For this study, we focus on Hill curves to model the conditional dose-responses. More detailed information can be found in Appendix A.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2.1.1 Loewe Additivity

Loewe Additivity builds on the concepts of sham combination and dose equivalence. The first concept is the idea that a compound does not interact with itself. The latter concept assumes that both compounds that reach the same effect can be interchanged. Therefore, any linear combination of fractions of those doses which reach the effect individually and, summed up, are equal to one, yields that exact same effect. Mathematically speaking, if dose x_1^* from the first compound reaches the same effect as dose x_2^* from the second compound, then any dose combination (x_1, x_2), for which

$$\frac{x_1}{x_1^*} + \frac{x_2}{x_2^*} = 1 \tag{3}$$

holds, should yield the same effect as x_1^* and x_2^*. As this idea can be generalized to any effect y, one gets

$$\frac{x_1}{f_1^{-1}(y)} + \frac{x_2}{f_2^{-1}(y)} = 1, \tag{4}$$

where x_1^* and x_2^* are replaced with $f_1^{-1}(y)$ and $f_2^{-1}(y)$, the inverse functions of Hill curves, respectively. For a fixed effect y, Eq. 4 defines an isobole, which is in mathematical terms a contour line. Hence the name of this model: the General Isobole Equation. It is an implicit formulation as the effect y of a dose combination (x_1, x_2) is implicitly given in Eq. 4. In the following we use the mathematical notation for the General Isobole Equation $f_{\mathrm{GI}}(x_1, x_2) = y$ with y being the solution to Eq. 4.

It was shown by Lederer et al. [2018a] that the principle of Loewe Additivity is based on a so-called Loewe Additivity Consistency Condition (LACC). This condition is that it should not matter whether equivalent doses of two compounds are expressed in terms of the first or the second. Under the assumption of the LACC being valid, Lederer et al. [2018a] have shown, that a null reference model can be formulated explicitly, by expressing the doses of one compound in terms of the other compound:

$$f_{2\to1}(x_1, x_2) = f_1\left(x_1 + f_1^{-1}(f_2(x_2))\right) \tag{5}$$

$$f_{1\to2}(x_1, x_2) = f_2\left(f_2^{-1}(f_1(x_1)) + x_2\right), \tag{6}$$

where $f_1^{-1}(f_2(x_2))$ is the dose x_1 of compound one to reach the same effect of compound two with dose x_2 (see Fig. 7 in Appendix A). For a detailed explanation, refer to Lederer et al. [2018a]. Summing up this dose equivalent of the first compound with the dose of the first compound allows for the computation of the expected effect of the compound combination. With the two formulations above, the effect y of the dose combination (x_1, x_2) is expressed as the effect of either one compound to reach that same effect. Under the LACC, all three models, Eq. 4, Eq. 5 and Eq. 6 are equivalent. It was further shown, that, in order for the LACC to hold, conditional dose-response curves must be proportional to each other, i.e. being parallel shifted on the x-axis in log-space. It has been commented by Geary [2012] and shown in [Lederer et al., 2018a], that this consistency condition is often violated. In an effort to take advantage of the explicit formulation and to counteract the different behavior of Eq. 5 and Eq. 6 in case of a violated LACC, Lederer et al. [2018a] introduced the so-called

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Explicit Mean Equation as mean of the two explicit formulations of Eq. 5 and Eq. 6:

$$f_{\text{mean}}(x_1, x_2) = 1/2\left(f_{2\to1}(x_1, x_2) + f_{1\to2}(x_1, x_2)\right). \tag{7}$$

A more extensive overview of Loewe Additivity and definition of null reference models together with visualizations can be found in Lederer et al. [2018a].

2.1.2 Bliss Independence

Bliss Independence assumes independent sites of action of the two compounds and was introduced a decade later than Loewe Additivity in [Bliss, 1939]. Note that the formulation of Bliss Independence depends on the measurement of the effect. The best known formulation of Bliss Independence is based on monotonically increasing responses for increasing doses:

$$g_{\text{bliss}}(x_1, x_2) = g_1(x_1) + g_2(x_2) - g_1(x_1)\, g_2(x_2), \tag{8}$$

where $g_i(x_i) = 1 - f_i(x_i)$ is a conditional response curve with increasing effect for increasing doses. In case the effect is measured in percent, i.e. $y \in [0, 100]$, the interaction term needs to be divided by 100 to ensure the right dimensionality of the term.

Here, we measure the effect in terms of cell survival or growth inhibition. Therefore the conditional response curves are monotonically decreasing for increasing concentrations or doses.

$$f_{\text{bliss}}(x_1, x_2) = f_1(x_1)\, f_2(x_2). \tag{9}$$

The records are normalized to the response at $x_1 = 0, x_2 = 0$, thus $f_1(0) = f_2(0) = 1$. To arrive from Eq. 8 to Eq. 9, one replaces any g by $1 - f$. Chou and Talalay [1984] derive the Bliss Independence from a first order Michaelis-Menten kinetic system with mutually non-exclusive inhibitors.

2.2 Methods

The models introduced in the previous section are null reference models in that they predict a response surface in the absence of compound interaction. We capture synergy in a single parameter to facilitate the screening process. This is different from other approaches, such as Chou and Talalay [1977], who measure synergy as deviation from a null-reference isobole without summarizing the deviation in a single parameter. The single parameter value is typically referred to as synergy- or α-score [Berenbaum, 1977]. As we investigate two methods to quantify synergy, we introduce two synergy parameters α and γ, which measure the extent of synergy. Both synergy scores α and γ are parametrized such that $\alpha = 0$ or $\gamma = 0$ denote absence of an interaction effect. In case α or γ take a value different from zero, we speak of a non-additive, or interactive effect. A compound combination is, dependent on the sign of synergy parameter, one of the three following:

$$\alpha, \gamma \begin{cases} > 0 & \text{synergistic} \\ = 0 & \text{additive or non-interactive} \\ < 0 & \text{antagonistic} \end{cases} \tag{10}$$

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Here, we measure synergy in two different ways, namely in fitting parametrized models or computing the lack-of-fit. The first method fits null reference models that are extended with a synergy parameter α. For these parametrized models α is computed by minimizing the square deviation between the measured response and the response spanned by the α-dependent model. For the second method the difference between a null reference model and the data is computed. For this method, the synergy score γ is defined as the volume that is spanned between the null reference model and the measured response.

Just as the conditional responses form the boundary condition for the null-reference surface (Eq. 1, Eq. 2), we want the conditional responses to be the boundary condition for all values of α. Explicitly, assuming a synergy model dependent on α is denoted by $f(x_1, x_2|\alpha)$, then

$$\left.\begin{array}{l} f(x_1, 0|\alpha) = f_1(x_1) \\ f(0, x_2|\alpha) = f_2(x_2) \end{array}\right\} \forall\alpha, \tag{11}$$

with f_i denoting the conditional response of compound i. We refer to Eq. 11 as the Synergy Desideratum. As we will see below, not all synergy models fulfill this property.

2.2.1 Parametrized Synergy

We extend the null reference models introduced in Section 2.1 in Eq. 4 - Eq. 9 to parametrized synergy models. The extension of the General Isobole Equation is the popular Combination Index introduced by Berenbaum [1977] and Chou and Talalay [1984]:

$$\frac{x_1}{f_1^{-1}(y)} + \frac{x_2}{f_2^{-1}(y)} = 1 - \alpha. \tag{12}$$

Berenbaum originally equated the left-hand side of Eq. 4 to the so-called Combination Index I. Depending on I smaller, larger, or equal to 1, synergy, antagonism or non-interaction is indicated. For consistency with the other synergy models, we set $I = 1 - \alpha$ such that α matches the outcomes as listed in Eq. 10. In Section 3 we will refer to this implicit model as $f_{\mathrm{CI}}(x_1, x_2|\alpha)$, where α is the parameter that minimizes the squared error between measured data and Eq. 12.

Note that this model violates the Synergy Desideratum in Eq. 11 as α not zero leads to deviations from the conditional responses. Explicitly, $f_{\mathrm{CI}}(x_1, 0|\alpha) = f_1((1-\alpha)x_1) \neq f_1(x_1)$. Although the Combination Index model violates the Synergy Desideratum, in practice it performs quite well and is in widespread use.

The explicit formulations in Eq. 5 and Eq. 6 are equivalent to the General Isobole Equation, $f_{\mathrm{GI}}(x_1, x_2)$, given in Eq. 4, under the LACC [Lederer et al., 2018a], but different if the conditional responses are not proportional. The two explicit equations are in fact an extension of the 'cooperative effect synergy' proposed by Geary [2012] for compounds with qualitatively similar effects. For these explicit formulations in Eq. 5 and Eq. 6 we propose a model that captures the interaction based on the explicit formulations:

$$f_{2\to1}(x_1, x_2|\alpha) = f_1\left(x_1 + (1+\alpha) f_1^{-1}(f_2(x_2))\right) \tag{13}$$

$$f_{1\to2}(x_1, x_2|\alpha) = f_2\left((1+\alpha) f_2^{-1}(f_1(x_1)) + x_2\right). \tag{14}$$

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

With this, we can extend the Explicit Mean Equation model $f_{\text{mean}}(x_1, x_2)$ in Eq. 7 to a parametrized synergy model:

$$f_{\text{mean}}(x_1, x_2|\alpha) = 1/2\left(f_{2\to1}(x_1, x_2|\alpha) + f_{1\to2}(x_1, x_2|\alpha)\right), \tag{15}$$

which we refer to as $f_{\text{mean}}(x_1, x_2|\alpha)$. As $f_{2\to1}(x_1, x_2|\alpha)$ and $f_{1\to2}(x_1, x_2|\alpha)$ do not fulfill the Synergy Desideratum, $f_{\text{mean}}(x_1, x_2|\alpha)$ also does not fulfill it.

To investigate the difference between the two models $f_{2\to1}(x_1, x_2)$ (Eq. 5) and $f_{1\to2}(x_1, x_2)$ (Eq. 6) we treat compound one and two based on the difference in slopes in the conditional responses (for more detailed information on the different parameters in Hill curves, refer to Appendix A). Instead of speaking of the first and second compound, we speak of the smaller and larger one, referring to the order of steepness. Therefore, we use models Eq. 13 and Eq. 14, but categorize the compounds based on the slope parameter of their conditional response curves. This results in $f_{\text{large}\to\text{small}}(x_1, x_2|\alpha)$ and $f_{\text{small}\to\text{large}}(x_1, x_2|\alpha)$.

Another synergy model we introduce here and refer to as $f_{\text{geary}}(x_1, x_2|\alpha)$ is based on a comment of Geary [2012], hence the naming. The two explicit models $f_{2\to1}(x_1, x_2)$ and $f_{1\to2}(x_1, x_2)$ yield the same surface under the LACC but do rarely in practice. Therefore, it cannot be determined whether a response that lies between the two surfaces is synergistic or antagonistic and hence should be treated as non-interactive. Thus, if α from $f_{1\to2}(x_1, x_2|\alpha)$ and α from $f_{2\to1}(x_1, x_2|\alpha)$ are of equal sign, the synergy score of that model is computed as the mean of those two parameters. In case the two synergy parameters are of opposite sign, the synergy score is set to 0:

$$\alpha_{\text{geary}} = \begin{cases} \frac{1}{2}(\alpha_{1\to2} + \alpha_{2\to1}) & \text{if } \operatorname{sign}(\alpha_{1\to2}) = \operatorname{sign}(\alpha_{2\to1}) \\ 0 & \text{else} \end{cases}. \tag{16}$$

Next, to extend the null reference model following the principle of Bliss Independence, we extend Eq. 8 to

$$g_{\text{bliss}}(x_1, x_2|\alpha) = g_1(x_1) + g_2(x_2) - (1+\alpha)\, g_1(x_1)\, g_2(x_2). \tag{17}$$

The motivation for this model is that any interaction between the two compounds is caught in the interaction term of the two conditional responses. In case of no interaction, the synergy parameter $\alpha = 0$, which leads to $(1+\alpha) = 1$, and results in no deviance from the null reference model. As we use the formulation of Eq. 9 due to measuring the effect as survival, we reformulate Eq. 17 analogously as we did to get from Eq. 8 to Eq. 9: by replacing $g_i(x_i)$ with $1 - f_i(x_i)$. Hence, Eq. 17 takes the form:

$$f_{\text{bliss}}(x_1, x_2|\alpha) = f_1(x_1)\, f_2(x_2) + \alpha\,(1 - f_1(x_1))\,(1 - f_2(x_2)) \tag{18}$$

This model does satisfy the requirement of no influence of the synergy parameter on conditional doses: $f_{\text{bliss}}(x_1, 0|\alpha) = f_1(x_1)$ and $f_{\text{bliss}}(0, x_2|\alpha) = f_2(x_2)$ as $f_i(0) = 1$. In case of synergy, the interactive effect is expected to be larger, therefore, α being positive. If the compound combination has an antagonistic effect, the interaction term is expected to be negative. For extreme α, the parametric approach leads to responses outside of the range $0 \leq y \leq 1$, e.g. $f_{\text{bliss}}(x_1, x_2) \to -\infty$ if $\alpha \to -\infty$. The same holds for the formulations of Loewe Additivity. The implicit formulation becomes impossible to match and for the

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

explicit formulations, the dose expression within brackets of $f_{2\to1}(x_1, x_2|\alpha)$ becomes negative. Additionally, $\alpha > 1$ is not possible for $f_{\mathrm{CI}}(x_1, x_2|\alpha)$, as the left-hand side of Eq. 12 can not be negative. Such behavior is also known from other models, e.g. for the Greco flagship model for negative synergy scores [Greco et al., 1995, p. 365-366, and Fig. 26]. Hence, we will limit α to the range of -1 to 1.

Despite of the Synergy Desideratum being violated for the models that build up on the Loewe Additivity principle, there is no further effect on the model comparison presented in Section 3 as conditional doses are excluded when computing the synergy score (see Section 2.2.2 and Section 2.2.3).

2.2.2 Lack-of-Fit Synergy

The second method to measure synergy investigated here is to compute the lack-of-fit of the measured response of a combination of compounds to the response of a null reference model derived from the conditional responses. We refer to this synergy value as γ:

$$\gamma = \int_{\min(x_2>0)}^{\max(x_2)} \int_{\min(x_1>0)}^{\max(x_1)} \left(\hat{y}(x_1, x_2|\Theta) - y(x_1, x_2)\right) d\log(x_1) d\log(x_2), \tag{19}$$

with $\hat{y}(x_1, x_2|\Theta)$ the estimated effect with parameters Θ of the fitted conditional responses following any non-interactive model and y the measured effect. Note that $\hat{y}(\Theta)$ and y are dependent on the concentration combination (x_1, x_2). This method was used in the AstraZeneca DREAM challenge [Menden et al., 2018] with the General Isobole Equation as null reference model and can be found in [Di Veroli et al., 2016]. Computing the volume has the advantage of taking the experimental design into account in contrast to simply taking the mean deviance over all measurement points, which is independent of the relative positions of the measurements. We also used a synergy value calculated from the mean deviance and it clearly performed worse (data not shown). The synergy value varies for different dose transformations. For example, the computed null-reference surface (and hence the synergy value) will be different for the same experiment if a log-transformation is applied to the doses or not.

In all, we have introduced six null reference models, five of them building up on the concept of Loewe Additivity and one on Bliss Independence. We further have introduced two methods to compute synergy, the parametric one and the lack-of-fit method, where both synergy parameters α and γ are positive if the record is synergistic, negative, if antagonistic. This results in twelve synergy model-method combinations: the parametric ones, $f_{\mathrm{CI}}(x_1, x_2|\alpha)$ (Eq. 12), $f_{\mathrm{large}\to\mathrm{small}}(x_1, x_2|\alpha)$ and $f_{\mathrm{small}\to\mathrm{large}}(x_1, x_2|\alpha)$ (Eq. 13, Eq. 14, dependent on the slope parameters) together with their mean, $f_{\mathrm{mean}}(x_1, x_2|\alpha)$ (Eq. 15), the method of Geary and $f_{\mathrm{bliss}}(x_1, x_2|\alpha)$ (Eq. 17). For the lack-of-fit method, we take as the null reference: $f_{\mathrm{GI}}(x_1, x_2)$ (Eq. 4), $f_{\mathrm{large}\to\mathrm{small}}(x_1, x_2)$ and $f_{\mathrm{small}\to\mathrm{large}}(x_1, x_2)$ (Eq. 5, Eq. 6), with the Explicit Mean Equation, $f_{\mathrm{mean}}(x_1, x_2)$ (Eq. 7), the method of Geary (analogously to Eq. 16) and $f_{\mathrm{bliss}}(x_1, x_2)$ (Eq. 9).

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2.2.3 Fitting the Synergy Parameter

Before applying the two methods presented in Section 2.2.1 and Section 2.2.2, we normalize and clean the data from outliers. In a first step we normalize all records to the same value, y_0, the measured response at zero dose concentration from both compounds. Second, we discard outliers using the deviation from a spline approximation. Third, we fit both conditional responses of each record, namely the responses of each compound individually, to a pair of Hill curves (Eq. 21, Appendix A). We fit the response at zero dose concentration for both Hill curves. This gives the parameter set $\mathbf{\Theta} = \{y_0, y_{\infty,1}, y_{\infty,2}, e_1, e_2, s_1, s_2\}$ for each record. More details are given in Appendix B.

We apply the two different methods to calculate the synergy parameters α and γ to each record. First, for the parametrized synergy models, we apply a grid search for α, for $\alpha \in [-1, 1]$ with a step size of 0.01, minimizing the sum of squared errors. This gives the value of α for which the squared error between the i^{th} measured effect $y^{(i)}$ and i^{th} expected effect $\hat{y}\left(x_1^{(i)}, x_2^{(i)} | \alpha, \Theta\right)$ is minimal:

$$\min_{\alpha} \sum_{i=1, \text{ with } x_1^{(i)} \neq 0 \text{ and } x_2^{(i)} \neq 0}^{N} \left(\hat{y}\left(x_1^{(i)}, x_2^{(i)} | \alpha, \Theta\right) - y^{(i)}\right)^2. \tag{20}$$

Note that we exclude the conditional responses that we used to fit $\mathbf{\Theta}$ from the minimization. Second, we apply the lack-of-fit method from Di Veroli et al. [2016], where synergy is measured in terms of the integral difference in log space of measured response and surface spanned by the non-interactive models in Section 2.1, as given in Eq. 19. For the calculation of the integrals, we apply the trapezoidal rule [Press et al., 2007, Chapter 4]. In Fig. 1 we summarize the most important steps of the analysis for a synergistic example. In Appendix E, Fig. 11, the same is shown for an antagonistic record.

2.3 Material

To evaluate the two methods introduced in Section 2.2.1 and Section 2.2.2, we apply them to two datasets of compound combination screening for which a categorization into the three synergy cases is provided.

The Mathews Griner dataset is a cancer compound synergy study by Mathews Griner et al. [2014]. In a one-to-all experimental design, the compound ibrutinib was combined with 463 other compounds and administered to the cancer cell line TMD8 of which cell viability was measured. The dataset is published at `https://tripod.nih.gov/matrix-client/`. Each compound combination was measured for 5 different doses, decreasing from $125\mu M$ to $2.5\mu M$ in a four-fold dilution for each compound alongside their conditional effects, resulting in 36 different dose combinations. The categorization of this dataset comes from a study by Yadav et al. [2015], in which every record was categorized based on a visual inspection.

The Cokol dataset comes from a study about fungal cell growth of the yeast S. cerevisiae (strain By4741), where Cokol et al. [2011] categorized the dataset. In this study the influence on cell growth was measured when exposed to 33 different compounds that were combined with one another based on promising combinations chosen by the authors, resulting in 200 different drug-drug-cell combinations. With an individually measured maximal effect dose for every

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

compound, the doses administered decrease linearly in seven steps with the eight dose set to zero, resulting in an 8×8 factorial design.

Based on the longest arc length of an isobole that is compared to the expected longest linear isobole in a non-interactive scenario, where Loewe Additivity serves as null reference model, each record was given a score. In more detail, from the estimated surface of a record assuming no interaction, the longest contour line is measured in terms of its length and direction (convex or concave). A convex contour line leads to the categorization of a record as synergistic and the arc length of the longest contour line determines the strength of synergy. A concave contour line results in an antagonistic categorization with its extent being measured again as the length of the longest isobole. Thus the Cokol dataset not only comes with a classification but also with a synergy score similar to α or γ.

To our knowledge, these two datasets are the only high-throughput ones with a classification into the three synergy classes: antagonistic, non-interactive and synergistic. Both datasets are somewhat imbalanced because interactions are rare [Borisy et al., 2003, Zhang et al., 2007, Farha and Brown, 2010]. The distribution of the classification is listed in Table 1. We obtained both categorizations after personal communication with the authors Yadav et al. [2015] and Cokol et al. [2011]. For the purpose of comparing the synergy models, we consider these two classifications as ground truth.

3 Results

Using the two methods of computing the synergy score, the parametric one (Section 2.2.1) and the lack-of-fit one (Section 2.2.2), we compute synergy scores for all records of the two datasets introduced in Section 2.3.

3.1 Kendall rank correlation coefficient

Having computed the synergy scores α and γ from the two different methods as described in Section 2.2.3, we compute the Kendall rank correlation coefficient, which is also known as Kendall's tau coefficient and was originally proposed by Kendall [1938]. This coefficient computes the rank correlation between the data as originally categorized by Yadav et al. [2015] and Cokol et al. [2011] and the computed synergy scores resulting from the two methods introduced in Section 2.2.1 and Section 2.2.2. For the analysis, we rank synergistic records highest at rank 3, followed by non-interactive at rank 2 and antagonistic lowest at rank 1. Due to the many ties in rank, the Kendall rank correlation coefficient cannot take a value higher than 0.75 for Mathews Griner and 0.8 for Cokol, even if a perfect ranking was given. An overview of the Kendall rank correlation coefficients is given in Table 3 and Table 4 in Appendix D.

To compare the parametric and lack-of-fit methods, we plot the correlation values as a scatter plot per method (see Fig. 2) with the values from the parametric method plotted on the x-axis and those from the lack-of-fit method on the y-axis. Most of the points scatter in the upper left triangle, above the diagonal line. This shows that the lack-of-fit method outperforms the parametric method. This holds for all models applied to the Mathews Griner dataset and also for all models but $f_{\text{geary}}(x_1, x_2|\alpha)$ and $f_{\text{small}\to\text{large}}(x_1, x_2|\alpha)$ applied

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

to the Cokol dataset. For both datasets, the highest correlation scores result from those null reference models that are based on the Loewe Additivity principle. The Bliss null reference model performs worst for the Mathews Griner set for both methods. For the Cokol data it is the second worst model. To a certain extent this can be explained due to the classification of the Cokol dataset being based on the isobole length relative to non-interactive isoboles, which is a Loewe Additivity type analysis. As the categorization of the Mathews Griner dataset is based on visual inspection, we cannot explain the bad performance of $f_{\text{bliss}}(x_1, x_2)$ for that dataset. On both datasets, $f_{\text{GI}}(x_1, x_2)$, $f_{\text{large}\to\text{small}}(x_1, x_2)$ and $f_{\text{mean}}(x_1, x_2)$ perform best for the lack-of-fit method. For the Mathews Griner dataset, $f_{\text{large}\to\text{small}}(x_1, x_2)$ dominates marginally over the General Isobole Equation and Explicit Mean Equation model. For the Cokol dataset, the Explicit Mean Equation dominates for both methods.

3.2 Scattering of Synergy Scores

To further investigate the performance of the methods and null reference models, we plot the synergy scores of the best performing models based on the Kendall rank correlation coefficient analysis (Section 3.1, and an ROC analysis, which we describe in detail in Appendix C) for both datasets in Fig. 3, Fig. 4 and Fig. 5. In all figures, the overall correlation of the compared data is depicted together with the correlation per categorization. The coloring of the scores is based on the original categorization as antagonistic, non-interactive or synergistic as provided by Yadav et al. [2015] and Cokol et al. [2011].

In Fig. 3 the synergy scores computed with the lack-of-fit method are plotted against the original synergy scores from Cokol et al. [2011]. Applying the lack-of-fit method to the Bliss Independence model (Eq. 9) results in scores which are mainly above zero (Fig. 3, upper left). Further, it can be seen in the density plots along the y-axis in Fig. 3, upper left panel, and on the x-axis of Fig. 4, both panels in the first row and left panel in the middle row, that the synergy scores that are computed based on the principle of Bliss Independence cannot be easily separated by categorization, making it difficult to come up with a threshold to categorize a record into one of the three synergy categories (synergy, antagonism, non-interaction) given a synergy score.

For the other three models depicted in Fig. 3, that are based on the principle of Loewe Additivity, the synergy scores are more clearly separated. The computed scores of the synergistic records distribute nicely above zero in the upper right corner (categorized as synergistic and computed synergy scores above zero) as well as they scatter in the lower left corner for antagonistic cases. In all those three panels in Fig. 3 we see for the non-interactive records that the computed scores of those three models are both positive and negative ranging roughly between -0.1 and 0.1 symmetrically. Barely any of the computed synergy scores for antagonistic cases are positive. Therefore, the chances of a record being antagonistic if the synergy score is above zero are quite low as well as the risk of categorizing a record as antagonistic if it is synergistic.

We further looked in detail into dose combinations for which both the $f_{\text{GI}}(x_1, x_2)$ and $f_{\text{mean}}(x_1, x_2)$ yield positive synergy values for antagonistic cases and into dose combinations for which the $f_{\text{mean}}(x_1, x_2)$ model results in negative synergy values for records which are labeled as synergistic. In total we found four dose combinations. A visualization of the observed and expected responses based on

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the Explicit Mean Equation model is shown in Appendix E, Fig. 12. One of them is a compound combined with itself. Hence, per definition of the Loewe Additivity, no interaction is expected. From Fig. 12, one can see why this record was mis-categorized: for high dose combinations, a greater effect is found, which is not found for the conditional runs. Probably, the dose ranges are too small to show such effects. We looked at the conditional responses of the other three dose combinations and observed that for the originally antagonistic records (three out of four) one of the conditional responses exhibits small effects with the maximal response y_∞ being above 0.65 (comp. right panel of Fig. 6). That leads to the computed null-reference surface to be quite high and hence causes synergistic scores if any effects are measured that are smaller than $\max\left(y_\infty^{(1)}, y_\infty^{(2)}\right)$. We suspect that the dose concentrations are not well-sampled and larger maximal doses should have been administered.

We further looked up the three dose combinations (excluding the one where the compound is combined with itself) in the Connectivity Map [Subramanian et al., 2017, Lamb et al., 2006], which is one of the largest repositories of drug response studies. Of those, we could find three in the Connectivity Map. All of these dose combinations showed non-interactive effects on all cell lines they were tested on. The assays found in the Connectivity Map are run on cancer cell lines. The dose combinations investigated here are run on yeast. Hence, a full comparison cannot be made, but results are certainly suggestive that the compound combinations are non-interactive.

In Fig. 4 and Fig. 5, the computed scores from different null reference models are plotted against each other. We compare the implicit formulation (General Isobole Equation) to the Bliss Independence model and the two best performing models that are based on the explicit formulation of Loewe Additivity, $f_{\text{mean}}\left(x_1, x_2\right)$ and $f_{\text{large}\to\text{small}}\left(x_1, x_2\right)$. The coloring of the scores is based on the original categorization as antagonistic, non-interactive or synergistic as provided by Yadav et al. [2015] and Cokol et al. [2011].

In Fig. 4 the scores from the Mathews Griner dataset are plotted. In the two panels in the upper row and the left panel in the middle row Bliss Independence is compared to the other three null models that build up on the principles of Loewe Additivity. It is obvious, that the scores based on Bliss Independence are larger than those of Loewe Additivity and mainly above zero. This is due to the more conservative null reference surface as derived from Bliss Independence (see [Sinzger et al., 2019, Fig. 6]). The scores from models that are based on Loewe Additivity are very similar to each other, as they scatter along the diagonal (panels in middle right and lower row). It is difficult, though, to tell apart whether a record is synergistic or antagonistic, as non-interactive records scatter largely between -0.5 and 0.5. Only records with a computed score outside that range can be categorized as interactive. For the Cokol dataset, which serves as basis for Fig. 5, the scores can be better separated. Despite the scores being generally smaller than those from the Mathews Griner data, the records can be easier separated, when using a Loewe Additivity based model. Additionally, we see here the similarity between these additive models given their strong correlation (right panels in middle row and both panels in lower row). Further, the scores based on $f_{\text{large}\to\text{small}}\left(x_1, x_2\right)$ achieve higher values than those from the other two Loewe Additivity based models. This becomes obvious when comparing the null-reference surfaces of those three models, as depicted

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

in [Lederer et al., 2018a, Fig. 4]. The surface spanned by $f_{\text{large}\to\text{small}}(x_1, x_2)$ spans a surface above those surfaces spanned by Explicit Mean Equation or General Isobole Equation. Therefore, in synergistic cases where the measured effect is greater, and hence the response in cell survival smaller, the difference from the null-reference surface to $f_{\text{large}\to\text{small}}(x_1, x_2)$ is greater than to the other two models. We suspect the synergy models from the Cokol dataset to be better separable due to the experimental design of the dataset. All compounds were applied up to their known maximal effect dose. This was not the case for the Mathews Griner dataset, where all compounds were applied at the same fixed dose range.

All in all, the lack-of-fit method performs better for any model when applied to the Mathews Griner dataset and mostly better for the Cokol dataset, with the exception of the $f_{\text{small}\to\text{large}}(x_1, x_2)$ and Geary model. We suggest, that the lack-of-fit should be preferred over the parametric method, due to the overall performance on both datasets. When using the lack-of-fit method, the Explicit Mean Equation model performs either second best (Mathews Griner dataset), or best (Cokol dataset). The other two well performing models, the explicit $f_{\text{large}\to\text{small}}(x_1, x_2)$ or the original implicit formulation of Loewe Additivity, the General Isobole Equation, do not perform equally well on both datasets. To exclude any bias from these models for different datasets, the Explicit Mean Equation should be preferred.

4 Discussion

The rise of high-throughput methods in recent years allows for massive screening of compound combinations. With the increase of data, there is an urge to develop methods that allow for reliable filtering of promising combinations. Additionally, the recent success of a synergy study of in vivo mice by Grüner et al. [2016] underlines the fast development of possibilities to generate biological data. Therefore, it is all the more important to develop methods that are sound and easily applicable to high-throughput data.

In this study we use two datasets of compound combinations that come with a categorization into synergistic, non-interactive or antagonistic for each record.

Based on the fitted conditional responses, we compute the synergy scores of all records. We compare six models that build on the principles of Loewe Additivity and Bliss Independence. Those six models are used with two different methods to compute a synergy score for each record. The first method is a parametric approach and is motivated by the Combination Index introduced by Berenbaum [1977]. The second method quantifies the difference in volume between the expected response assuming no interaction and the measured response and is motivated by Di Veroli et al. [2016].

We compare the computed synergy scores from both methods, each applied with the six reference models, based on Kendall rank correlation coefficients. Based on these correlation coefficients we investigate the reconstruction of ranking of the records (see Section 3.1). We further conduct an ROC analysis (results shown in Appendix C). With this, we quantify the methods' and models' capacity to distinguish records from different categories, given a computed synergy score. Both, the Kendall rank correlation coefficient and the ROC analysis, show a superiority of those models that are based on Loewe Additivity relative

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

to those based on Bliss Independence. From those additive models the Explicit Mean Equation is the overall best performing model for both datasets.

For the above comparison of the six null reference models and the two methods, we rely on the underlying categorization of both datasets. All performance metrics are based on how well the predicted synergy scores agree with the underlying categorization. The categorizations of both datasets were created very differently from one another. On one hand, the Mathews Griner dataset was categorized on a visual inspection, on account of which we cannot be certain about the assumptions made that guided the decision making process. On the other hand, the categorization of the Cokol dataset is based on the principle of Loewe Additivity. This leads to the natural preference of null models that are based on Loewe Additivity over those based on Bliss Independence, which we find back in our analysis. Irrespective of the origin of the classification, we stress that the labels were provided to us by independent researchers and hence were not biased in any way to favor the Explicit Mean Equation model.

Note that we conduct the research only on combinations of two compounds. Meanwhile, it is shown in Russ and Kishony [2018] that Bliss Independence maintains accuracy when increasing the number of compounds that are combined with each other. Loewe Additivity, however, loses its predictive power for an increasing number of compounds.

The comparison of the parametric method with the lack-of-fit method shows a superiority of the lack-of-fit method. To recall, the motivation behind the parametric approach was the statistical advantages of such an approach. It allows to define an interval around $\alpha = 0$ in which a compound combination can be considered additive. For the lack-of-fit method, such statistical evaluation can not be done directly, but could be performed on the basis of bootstrapping.

Chou and Talalay [1977] measure the interaction effect locally for a fixed ratio of doses of both compounds that are supposed to reach the same effect, say one unit of the first compound causes the same effect as two units of the second compound, which results in the dose combination of 1:2. Along this fixed ratio of doses, they compute the left-hand side of Eq. 3 given the two doses x_1 and x_2 that are assumed to reach a fixed effect y^* together with x_i^* being the dose of compound i that reaches the fixed effect alone. For the fixed dose ratio, they run over all expected effects, usually from zero to one. A geometric interpretation of that method is depicted in [Greco et al., 1995, Fig. 7, p. 341]. The resulting values of the left-hand side of Eq. 3 are analyzed graphically: all computed values are plotted versus the expected fixed effect $y^* = [0, 1]$. Values higher than one exhibit synergistic behavior, values below one antagonism. This method allows for results that show antagonistic behavior for, say, smaller effects, as well as synergistic behavior for higher effects, or vice versa. That such a behavior of switching from antagonistic behavior in one region to synergistic behavior in another can occur was also shown in Norberg and Wahlström [1988]. With one synergy score, as used throughout this paper, we do not provide such a measure for local antagonism and synergism. Our main motivation in this study is to provide a single synergy score that allows for fast filtering of interesting candidates for more in-depth research. To extend that idea, the standard deviation could be taken into account, as in a t-value or Z-score. Additionally, the superior lack-of-fit method is much faster and simpler to implement than the parametric one.

Finally, to asses how distinguishable the synergy scores γ are, we visualize

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the synergy scores based on the underlying category (Section 3.2). The synergy scores from the lack-of-fit method can, based on their sign, reliably be categorized as synergistic or antagonistic. For records categorized as non-interactive, the computed synergy scores are positive as well as negative. For the two datasets, we saw different extents of separation between those γ-scores, which makes it difficult to generalize the results. All in all, the differentiation from no interaction poses a more difficult task as choosing the threshold is arbitrary.

During the analysis, we observed higher synergy scores when applying the Bliss Independence principle as null reference model. This is due to the more conservative null reference surface as derived from Bliss Independence (see exemplary comparison of isoboles from most of the models discussed here in [Sinzger et al., 2019, Fig. 6]). Due to the synergy scores being relatively high, a differentiation between categories based on the synergy score poses a bigger challenge. There is a strong overlap of synergy scores from all three categories. Additionally, most of the synergy scores γ, that are computed with the lack-of-fit method, are above zero. Different ranges of synergy scores for both datasets make it additionally difficult to assess synergy or antagonism for a record based on the unique information of the synergy score.

We want to emphasize the performance benefit of the recently introduced Explicit Mean Equation [Lederer et al., 2018a] over the implicit formulation in form of the General Isobole Equation. On both datasets, it is the overall best performing model when compared to the provided categorizations. The explicit formulation of this additive model was shown to speed up computation by a factor of 250 (see [Lederer et al., 2018a, Fig. S1]). Together with the implementation of the lack-of-fit method, which is easier to implement and a lot faster than the parametric method, this combination of model improvement and method can be of great benefit for the research community.

Although the performance of models and methods are consistent across the two (quite different) datasets considered in this study, reliable comparison of different models and methods would benefit from the availability of drug screening datasets that are available with ground truth labeling.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Acknowledgments

We thank Bhagwan Yadav for the sharing of the code used for the analysis in Yadav et al. [2015] and Murat Cokol for the sharing of the data and analytical insights from Cokol et al. [2011].

An earlier version of this manuscript has been released as a Pre-Print at [Lederer et al., 2018b].

Conflict of Interest

The authors declare that they have no conflict of interest.

Funding

This work was supported by the Radboud University and CogIMon H2020 ICT-644727.

References

S Loewe. Die quantitativen Probleme der Pharmakologie. *Ergebnisse der Physiol.*, 27(1):47–187, 1928. ISSN 03034240. doi: 10.1007/BF02322290.

C I Bliss. The toxicity of poisons applied jointly. *Ann. Appl. Biol.*, 26(3): 585–615, 1939. ISSN 00034746. doi: 10.1111/j.1744-7348.1939.tb06990.x.

Simone Lederer, Tjeerd M H Dijkstra, and Tom Heskes. Additive Dose Response Models: Explicit Formulation and the Loewe Additivity Consistency Condition. *Front. Pharmacol.*, 9(February):1–11, 2018a. doi: 10.3389/fphar.2018.00031. URL `https://www.frontiersin.org/articles/10.3389/fphar.2018.00031/full`.

Bhagwan Yadav, Krister Wennerberg, Tero Aittokallio, and Jing Tang. Searching for Drug Synergy in Complex Dose-Response Landscapes Using an Interaction Potency Model. *Comput. Struct. Biotechnol. J.*, 13:504–513, 2015. ISSN 20010370. doi: 10.1016/j.csbj.2015.09.001. URL `http://dx.doi.org/10.1016/j.csbj.2015.09.001`.

Murat Cokol, Hon Nian Chua, Murat Tasan, Beste Mutlu, Zohar B Weinstein, Yo Suzuki, Mehmet E Nergiz, Michael Costanzo, Anastasia Baryshnikova, Guri Giaever, Corey Nislow, Chad L Myers, Brenda J Andrews, Charles Boone, and Frederick P Roth. Systematic exploration of synergistic drug pairs. *Mol. Syst. Biol.*, 7(544), 2011. ISSN 1744-4292. doi: 10.1038/msb.2011.71.

William R Greco, Gergory Bravo, and John C Parsons. The Search for Synergy: A Critical Review from A Response Surface Perspective. *Pharmacol. Rev.*, 47(2):331–385, 1995.

Nori Geary. Understanding Synergy. *Am. J. Physiol. Endocrinol. Metab.*, 304 (3):E237–E253, 2012. ISSN 1522-1555. doi: 10.1152/ajpendo.00308.2012. URL `http://www.ncbi.nlm.nih.gov/pubmed/23211518`.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Charles F Minto, Thomas W Schnider, Timothy G Short, Keith M Gregg, Andrea Gentilini, and Steven L Shafer. Response Surface Model for Anesthetic Drug Interactions. *Anestesiology*, 92(6):1603–1606, 2000.

Ting-Chao Chou and Paul Talalay. Quantitative analysis of dose-effect relationships: the combined effects of multiple drugs or enzyme inhibitors. *Adv. Enzyme Regul.*, 22:27–55, 1984. doi: 10.1016/0065-2571(84)90007-4.

Giovanni Y Di Veroli, Chiara Fornari, Dennis Wang, Severine Mollard, Jo L Bramhall, Frances M Richards, and Duncan I Jodrell. Combenefit: An interactive platform for the analysis and visualisation of drug combinations. *Bioinformatics*, 32(18):2866–2868, 2016. ISSN 1367-4803. doi: 10.1093/bioinformatics/btw230. URL `http://bioinformatics.oxfordjournals.org/content/early/2016/05/27/bioinformatics.btw230.full.pdf{%}5Cnhttp://bioinformatics.oxfordjournals.org/content/early/2016/04/25/bioinformatics.btw230.abstract`.

D. J. Hand. Synergy in drug combinations. In Wolfgang Gaul, Otto Opitz, and Martin Schader, editors, *Data Anal. Sci. Model. Pract. Appl.*, pages 471–475. Springer-Verlag, Berlin Heidelberg, 2000. ISBN 978-3-540-67731-4.

Mark Sinzger, Jakob Vanhoefer, Carolin Loos, and Jan Hasenauer. Comparison of null models for combination drug therapy reveals Hand model as biochemically most plausible. *Sci. Rep.*, 9(3002):1–15, 2019. ISSN 2045-2322. doi: 10.1038/s41598-019-38907-x. URL `http://dx.doi.org/10.1038/s41598-019-38907-x`.

Liye He and Jing Tang. *synergyfinder: Calculate and Visualize Synergy Scores for Drug Combinations*, 2016. URL `https://cran.r-project.org/package=synergyfinder`.

Christian T Meyer, David J Wooten, B Bishal Paudel, Joshua Bauer, Keisha N Hardeman, David Westover, Christine M Lovly, Leonard A Harris, Darren R Tyson, and Vito Quaranta. Quantifying Drug Combination Synergy along Potency and Efficacy Axes. *Cell Syst.*, 8(2):97–108, 2019. ISSN 2405-4712. doi: 10.1016/j.cels.2019.01.003. URL `https://doi.org/10.1016/j.cels.2019.01.003`.

Ronald J W Lambert and Douglas A Dawson. individual and combined toxicants. *Toxicol. Res. (Camb).*, 2019. doi: 10.1039/c9tx00005d.

Powerful combination therapies. *Nat. Biomed. Eng.*, 2(8):555–556, 2018. ISSN 2157-846X. doi: 10.1038/s41551-018-0283-1. URL `https://doi.org/10.1038/s41551-018-0283-1`.

Joseph Lehar, Grant R Zimmermann, Andrew S Krueger, Raymond A Molnar, Jebediah T Ledell, Adrian M Heilbut, Glenn F Short, Leanne C Giusti, Garry P Nolan, Omar A Magid, Margaret S Lee, Alexis A Borisy, Brent R Stockwell, and Curtis T Keith. Chemical combination effects predict connectivity in biological systems. *Mol. Syst. Biol.*, 3(80):80, 2007. ISSN 1744-4292. doi: 10.1038/msb4100116.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Ting-Chao Chou and Paul Talalay. A Simple Generalized Equation for the Analysis of Multiple Inhibitions of Michaelis-Menten Kinetic Systems. *J. Biol. Chem.*, 252(18):6438–6442, 1977.

M.C. Berenbaum. Synergy, additivism and antagonism in immunosuppression. *Clin. exp. Immunol.*, 28:1–18, 1977.

Michael Patrick Menden, Dennis Wang, Yuanfang Guan, Michael Mason, Bence Szalai, Krishna C Bulusu, Thomas Yu, Jaewoo Kang, Minji Jeon, Russ Wolfinger, Tin Nguyen, Mikhail Zaslavskiy, In Sock Jang, Zara Ghazoui, Mehmet Eren Ahsen, Robert Vogel, Elias Chaibub Neto, Thea Norman, Eric K Y Tang, Mathew J Garnett, Giovanni Di Veroli, Stephen Fawell, Gustavo Stolovitzky, Justin Guinney, Jonathan R Dry, and Julio Saez-Rodriguez. A cancer pharmacogenomic screen powering crowd-sourced advancement of drug combination prediction. *bioRxiv*, 2018. URL `http://biorxiv.org/content/early/2018/02/13/200451.abstract`.

William H Press, Saul A Teukolsky, William T Vetterling, and Brian P Flannery. *Numerical Recipes: The Art of Scientific Computing*, volume 1. Cambridge University Press, Cambridge, 3 edition, 2007. ISBN 0521880688. doi: 10.1137/1031025.

Lesley A Mathews Griner, Rajarshi Guha, Paul Shinn, Ryan M Young, Jonathan M Keller, Dongbo Liu, Ian S Goldlust, Adam Yasgar, Crystal McKnight, Matthew B Boxer, Damien Y Duveau, Jian-Kang Jiang, Sam Michael, Tim Mierzwa, Wenwei Huang, Martin J Walsh, Bryan T Mott, Paresma Patel, William Leister, David J Maloney, Christopher A Leclair, Ganesha Rai, Ajit Jadhav, Brian D Peyser, Christopher P Austin, Scott E Martin, Anton Simeonov, Marc Ferrer, Louis M Staudt, and Craig J Thomas. High-throughput combinatorial screening identifies drugs that cooperate with ibrutinib to kill activated B-cell-like diffuse large B-cell lymphoma cells. *Proc. Natl. Acad. Sci. U. S. A.*, 111(6):2349–54, 2014. ISSN 1091-6490. doi: 10.1073/pnas.1311846111. URL `http://www.pnas.org/cgi/content/long/111/6/2349`.

Alexis A Borisy, Peter J Elliott, Nicole W Hurst, Margaret S Lee, Joseph Lehar, E Roydon Price, George Serbedzija, Grant R Zimmermann, Michael A Foley, Brent R Stockwell, and Curtis T Keith. Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci. U. S. A.*, 100(13):7977–7982, 2003. ISSN 0027-8424. doi: 10.1073/pnas.1337088100.

L Zhang, K Yan, Y Zhang, R Huang, J Bian, C Zheng, H Sun, Z Chen, N Sun, R An, F Min, W Zhao, Y Zhuo, J You, Y Song, Z Yu, Z Liu, K Yang, H Gao, H Dai, X Zhang, J Wang, C Fu, G Pei, J Liu, S Zhang, M Goodfellow, Y Jiang, J Kuai, G Zhou, and X Chen. High-throughput synergy screening identifies microbial metabolites as combination agents for the treatment of fungal infections. *Proc. Natl. Acad. Sci. U. S. A.*, 104(11):4606–4611, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0609370104. URL `http://www.ncbi.nlm.nih.gov/pubmed/17360571`.

Maya A Farha and Eric D Brown. Chemical probes of escherichia coli uncovered through chemical-chemical interaction profiling with compounds of known biological activity. *Chem. Biol.*, 17(8):852–862, 2010. ISSN 10745521.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

doi: 10.1016/j.chembiol.2010.06.008. URL http://dx.doi.org/10.1016/j.chembiol.2010.06.008.

M G Kendall. A new Measure of Rank Correlation. *Biometrika*, 30(1-2):81–93, 1938. doi: 10.1093/biomet/30.1-2.81. URL http://dx.doi.org/10.1093/biomet/30.1-2.81.

Aravind Subramanian, Rajiv Narayan, Steven M Corsello, David D Peck, Ted E Natoli, Xiaodong Lu, Joshua Gould, John F Davis, Andrew A Tubelli, Jacob K Asiedu, David L Lahr, Jodi E Hirschman, Zihan Liu, Melanie Donahue, Bina Julian, Mariya Khan, David Wadden, Ian C Smith, Daniel Lam, Arthur Liberzon, Courtney Toder, Mukta Bagul, Marek Orzechowski, Oana M Enache, Federica Piccioni, Sarah A Johnson, Nicholas J Lyons, Alice H Berger, Alykhan F Shamji, Angela N Brooks, Anita Vrcic, Corey Flynn, Jacqueline Rosains, David Y Takeda, Roger Hu, Desiree Davison, Justin Lamb, Kristin Ardlie, Larson Hogstrom, Peyton Greenside, Nathanael S Gray, Paul A Clemons, Serena Silver, Xiaoyun Wu, Wen Ning Zhao, Willis Read-Button, Xiaohua Wu, Stephen J Haggarty, Lucienne V Ronco, Jesse S Boehm, Stuart L Schreiber, John G Doench, Joshua A Bittker, David E Root, Bang Wong, and Todd R Golub. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, 171(6): 1437–1452.e17, 2017. ISSN 10974172. doi: 10.1016/j.cell.2017.10.049. URL http://dx.doi.org/10.1016/j.cell.2017.10.049.

Justin Lamb, Emily D Crawford, David Peck, Joshua W Modell, Irene C Blat, Matthew J Wrobel, Jim Lerner, Jean-Philippe Brunet, Aravind Subramanian, Kenneth N Ross, Michael Reich, Haley Hieronymus, Guo Wei, Scott A Armstrong, Stephen J Haggarty, Paul A Clemons, Ru Wei, and Steven A Carr. The Connectivity Map : Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science (80-.).*, 313(September):1929–1935, 2006. ISSN 1095-9203. doi: 10.1126/science.1132939.

Barbara M Grüner, Christopher J Schulze, Dian Yang, Daisuke Ogasawara, Melissa M Dix, Zoë N Rogers, Chen-Hua Chuang, Christopher D McFarland, Shin-Heng Chiou, J Mark Brown, Benjamin F Cravatt, Matthew Bogyo, and Monte M Winslow. An in vivo multiplexed small-molecule screening platform. *Nat. Methods*, 13(September), 2016. ISSN 1548-7091. doi: 10.1038/nmeth.3992. URL http://www.nature.com/doifinder/10.1038/nmeth.3992.

D Russ and R Kishony. Additivity of inhibitory effects in multidrug combinations. *Nat. Microbiol.*, 2018. ISSN 2058-5276. doi: 10.1038/s41564-018-0252-1. URL http://dx.doi.org/10.1038/s41564-018-0252-1.

L Norberg and G Wahlström. Anaesthetic effects of flurazepam alone and in combination with thiopental or hexobarbital evaluated with an EEG-threshold method in male rats. *Arch. Int. Pharmacodyn. Ther.*, 292:45–57, 1988.

Simone Lederer, Tjeerd M H Dijkstra, and Tom Heskes. Additive Dose Response Models: Defining Synergy. *bioRxiv*, 2018b. doi: 10.1101/480608. URL https://www.biorxiv.org/content/early/2018/11/29/480608.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A V Hill. The possible effects of the aggregation of the molecule of hemoglobin on its dissociation curves. *J. Physiol.*, 40:iv–vii, 1910.

Sylvain Goutelle, Michel Maurin, Florent Rougier, Xavier Barbaut, Laurent Bourguignon, Michel Ducher, and Pascal Maire. The Hill equation: A review of its capabilities in pharmacological modelling. *Fundam. Clin. Pharmacol.*, 22(6):633–648, 2008. ISSN 07673981. doi: 10.1111/j.1472-8206.2008.00633.x.

Christian Ritz and Jens C Strebig. *drc: Analysis of Dose-Response Curves*, 2016. URL `https://cran.r-project.org/package=drc`.

Simon Wood. *Generalized Additive Models: an introduction with R*. Chapman and Hall/CRC, 2 edition, 2017. ISBN 1584884746. doi: 10.1111/j.1541-0420.2006.00574.x.

Simon N Wood. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. R. Stat. Soc.*, 73(1):3–36, 2011. ISSN 13697412. doi: 10.1111/j.1467-9868.2010.00749.x.

Christian Ritz, Florent Baty, Jens C Streibig, and Daniel Gerhard. Dose-response analysis using R. *PLoS One*, 10(12):1–13, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0146021. URL `http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0146021`.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016. URL `https://www.r-project.org/`.

Takaya Saito and Marc Rehmsmeier. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLoS One*, 10(3):1–21, 2015. ISSN 19326203. doi: 10.1371/journal.pone.0118432.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Tables

	synergistic	antagonistic	non-interactive
Mathews Griner	121	90	252
Cokol	50	68	82

Table 1: Number of cases categorized as synergistic, antagonistic or non-interactive in the two datasets Mathews Griner and Cokol.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure Captions

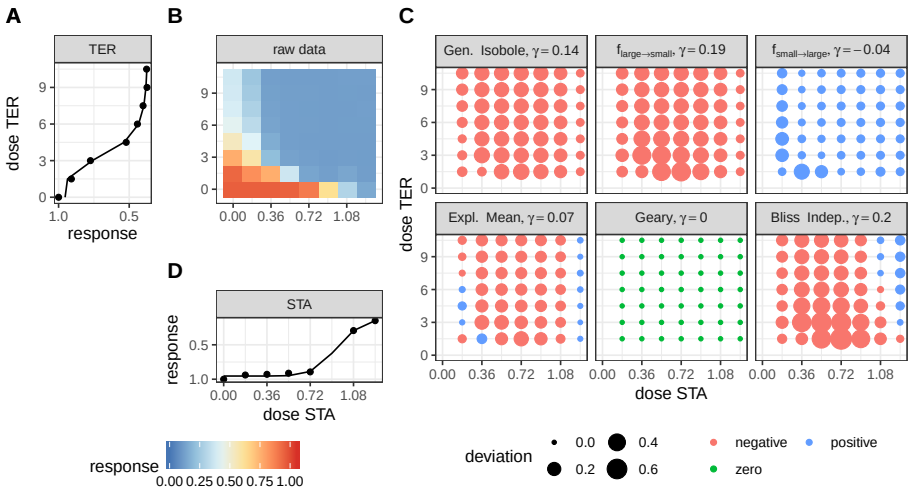

Figure 1: Description of the analysis steps of the lack-of-fit method for the compound pair TER and STA from the Cokol dataset. This compound pair is categorized as synergistic according to [Cokol et al., 2011]. The raw response data of the record is depicted in (B). The response data normalized by the read at zero dose concentration (lower left). In (B) the degree of relative cell growth is colored from high to low values in red to blue.
Step 1: compute Hill curves for conditional responses: Based on the raw reads of the single dose responses (lower and left outer edges) fit a Hill curve to the conditional responses. The fitted Hill curves shown in (A) and (D) with original raw data shown as points.
Step 2: compute expected non-interactive response for all six models: not shown.
Step 3: compute difference between measured data (C) and expected data from all six null reference models: shown in (C). The direction of difference is shown by color (red for negative and blue for positive, green for zero). The larger the degree of difference, the larger the bullet, and vice versa.
Step 4: compute integral γ over the differences: Over all those bullets, we then compute the integral, which gives the synergy score γ. For every model, the synergy score γ is depicted in the title of each matrix in (C).

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

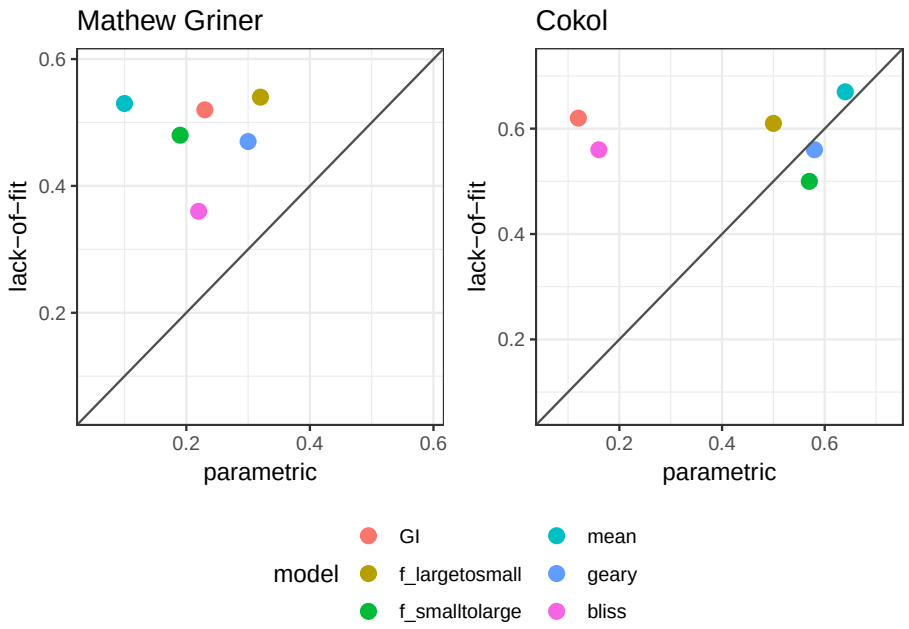

Figure 2: Scatter plot of Kendall rank correlation coefficient for both datasets, Mathews Griner (left) and Cokol (right). The Kendall correlation measures the rank correlation of the original categorization and the computed synergy scores. The higher the correlation, the more similar the score ranking. The correlation values from the synergy scores α, computed with the parametric approach, are plotted on the x-axis and those from the lack-of-fit approach are plotted on the y-axis. Each model is depicted in a different color. To guide the eye, the diagonal is plotted. If a data point is above the diagonal, the Kendall rank correlation coefficient from the lack-of-fit method is higher than that from the parametric method, and vice versa. Without exception, the Kendall rank correlation coefficients are all higher for the synergy scores γ, which are computed with the lack-of-fit method, than those based on the α scores computed with the parametric method.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

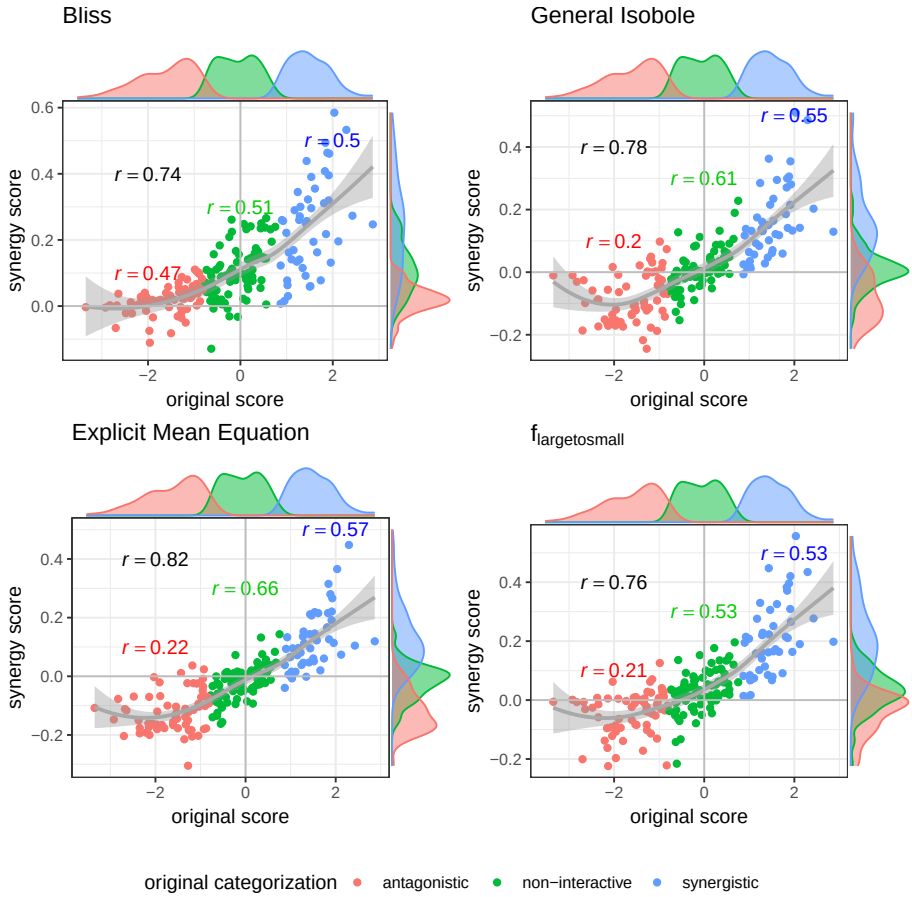

Figure 3: Computed synergy scores γ of Cokol data of the best models according to the Kendall rank correlation coefficient and ROC analysis in Section 3.1 and Appendix C in comparison to the original scores from Cokol et al. [2011]. The data points are colored based on the original categorization. For all three categories, synergistic, non-interactive and antagonistic, the Pearson correlation is depicted between the original scores in that category and the computed synergy scores in the respective color. Additionally, we depict the local polynomial regression fitting of all scores (in gray). The histograms of the scores are plotted on the axis, separated by color based on the original categorization. Synergy scores γ based on the Explicit Mean Equation model show the highest correlation with the original scores.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

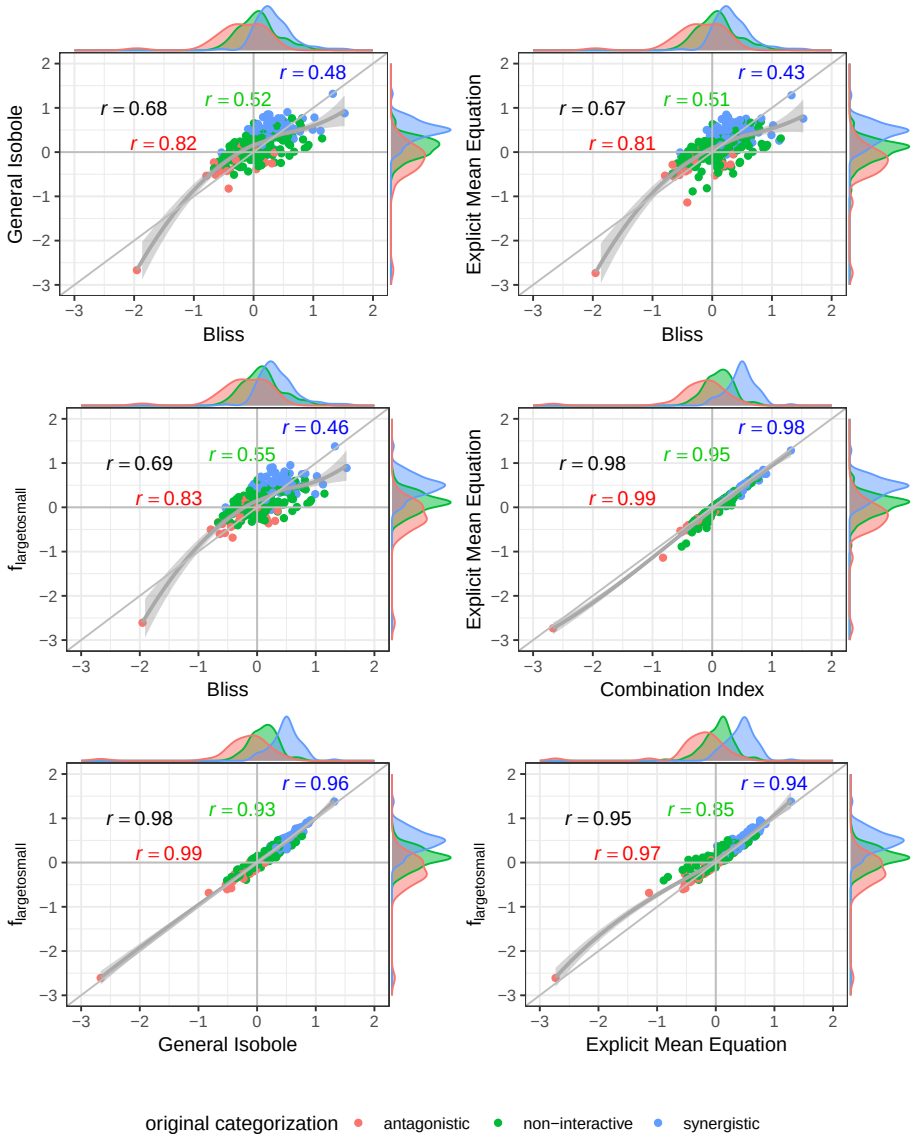

Figure 4: Scatter plot of synergy scores γ of the Mathews Griner dataset. The scores are computed with the lack-of-fit method. Displayed are the four best models according to the Kendall rank correlation coefficient and ROC analysis in Section 3.1 and Appendix C. The scores of one model are depicted on the x-axis and the other on the y-axis. The original categorization is given based on colour. The Pearson correlation score between the synergy scores are depicted by color for every categorization and the overall Pearson correlation is depicted in black. To guide the eye, the axis at 0, the diagonal and local polynomial regression fitting are depicted in grey. The histograms of the scores are plotted on the axis, separated by color based on the original categorization. The three models based on the Loewe Additivity principle show highest correlation (center right and lower row). All comparison with $f_{\text{bliss}}(x_1, x_2)$ show lowest correlation (first three cases). There is a large difference between the correlation between the additive models and the comparison of Bliss Independence by roughly 0.3.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

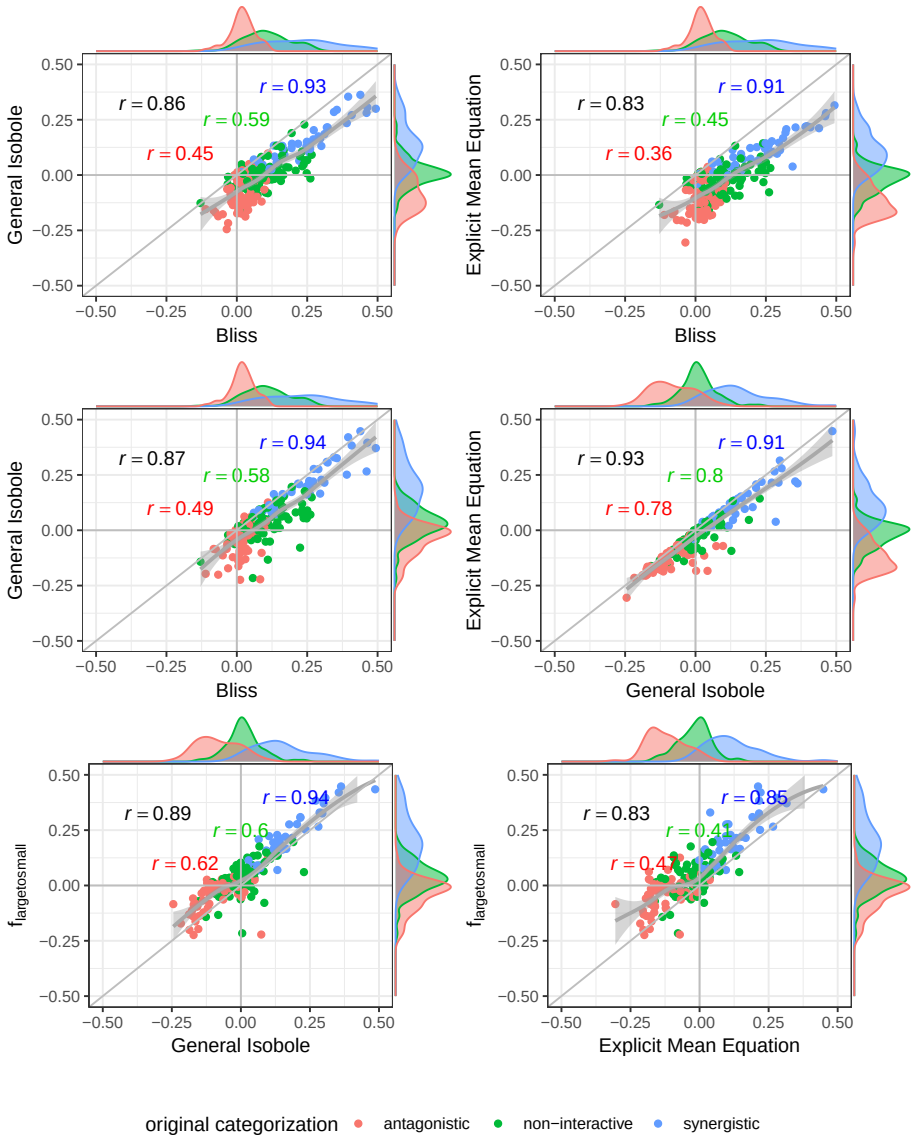

Figure 5: Scatter plot of synergy scores γ of the Cokol dataset. The scores are computed with the lack-of-fit method. Displayed are the four best models according to the Kendall rank correlation coefficient and ROC analysis in Section 3.1 and Appendix C. The scores of one model are depicted on the x-axis and the other on the y-axis. The original categorization is given based on colour. The Pearson correlation score between the synergy scores are depicted by color for every categorization and the overall Pearson correlation is depicted in black. To guide the eye, the axis at 0, the diagonal and the local polynomial regression fitting are depicted in grey. The histograms of the scores are plotted on the axis, separated by color based on the original categorization. $f_{\text{mean}}(x_1, x_2)$ and $f_{\text{GI}}(x_1, x_2)$ show highest correlation (center right), $f_{\text{bliss}}(x_1, x_2)$ shows lowest (first three comparison cases).

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

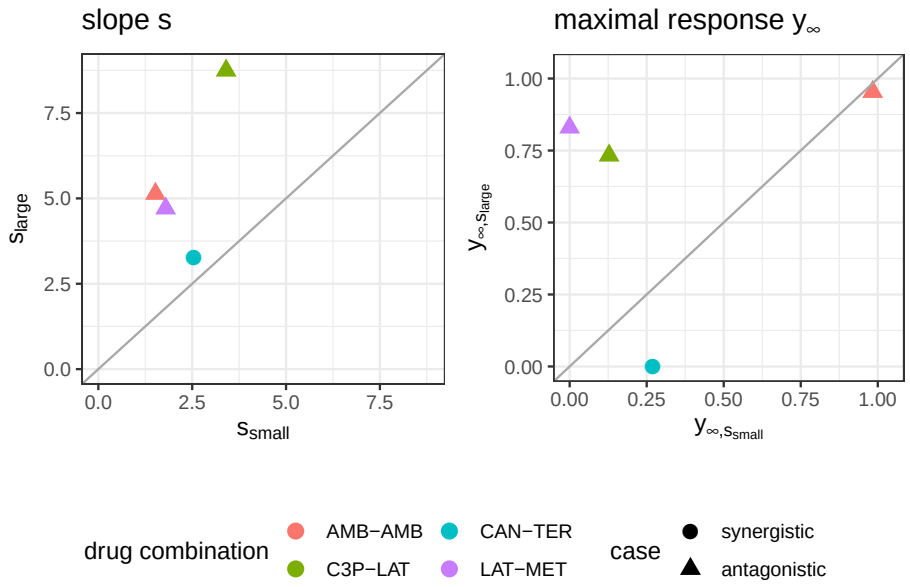

Figure 6: Maximal response y_∞ (left) and slope parameters s (right) of Hill curves. Parameters are shown for the conditional responses of the four cases for which the lack-of-fit method resulted for $f_{\text{mean}}(x_1, x_2)$ and $f_{\text{GI}}(x_1, x_2)$ in a synergy score of opposite sign to its categorization from the Cokol dataset. Different records are depicted in different colours. The original categorization of each record is depicted per shape. The conditional responses of one record, and hence their Hill curve parameters, are grouped depending on size of the Hill curve parameter s (larger or smaller).

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A Conditional Dose Response Curves

A common approach for modeling monotonic dose-response curves f_j with $j \in \{1, 2\}$ is the Hill curve [Hill, 1910], also referred to as the sigmoid function. The Hill model is, due to its good fit to many sources of data, the most widely applied model for fitting compound responses [Goutelle et al., 2008]. It has a sigmoidal shape with little change for small doses but with a rapid decline in response once a certain threshold is met. For even larger doses the effect asymptotes to a constant maximal effect. Two exemplary Hill curves are depicted in Fig. 7. There are several parameterizations of the Hill curve. We use the following

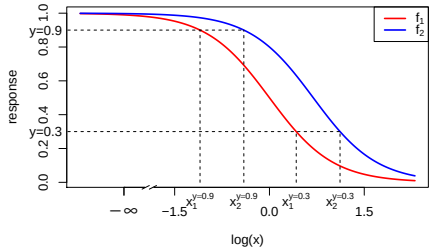

Figure 7: Dose-response curves (red and blue) as Hill curves (Eq. 21). For the exemplary responses of 0.3 and 0.9 the different doses x_1 and x_2 reaching that effect are shown (dashed lines). The dose-response curves differ only in EC_{50} with $e_1 = 2$ and $e_2 = 1$. Values of the other parameters are $y_0 = 1$, $y_\infty = 0$ and $s = 2$. To highlight the sigmoidal shape of a Hill curve in log-space, the logarithmic concentration space is depicted.

throughout this study to fit conditional responses:

$$f(x) = y_\infty + \frac{y_0 - y_\infty}{1 + (\frac{x}{e})^s}, \tag{21}$$

where y_0 is the response at zero dose and y_∞ the maximal response of the cells to the compound, e the dose concentration reaching half of the maximal response and s the steepness of the curve. Eq. 21 is equivalent to the parametrization used in the drc package [Ritz and Strebig, 2016], the so-called four parameter log-logistic model. By our definition of the Hill curve, a positive s leads to a descending Hill curve.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

B Data Cleaning, Fitting of Hill Curve and Parameter Estimation for Implicit Models

First, we normalize all records by the measured response at zero dose concentration from both compounds, y_0. Second, we conduct an outlier analysis of the normalized responses by fitting a spline surface and deleting outliers to discard them. Third, we then fit the conditional responses of the cleaned data to Hill curves.

We fit a general additive model (GAM) to the normalized raw data using thin plate splines [Wood, 2017], not transforming the doses in any way. The surfaces of those fitted thin plate splines span the checkerboards of every record and data points with too large absolute residual values are rejected. For fitting the splines we use method gam() of the mgcv-package [Wood, 2011], defining the smooth terms within the gam formulae with the method s(). We set the dimension of the basis, that is used to represent the smooth term to $k = 30$ fixed knots.

The threshold to reject data points is at five times the inter-quantile range of all residuals of a given record. Every data point with an absolute residual above that threshold is discarded. For the Mathews Griner data, this leads to 18 records out of the 466 (less than 4%) where a mean of 1.28 outliers were excluded per record with an overall of 23 data points excluded, which is less than one percent of the overall data. A maximum of 6 outliers was detected once. Similarly, we excluded on average 2.48 data points for the Cokol data on 52 of the total 200 ($\leq 25\%$) records with a maximum of 13 data points and an overall of 129 data points excluded, which is about 1% of all data points.

To fit the two conditional responses of a record to two Hill functions of the form of Eq. 21 we use the drc package [Ritz et al., 2015]. Unlike other synergy analyses such as [Yadav et al., 2015], the response at zero concentration y_0 is not fixed to 1 but merely constrained to be the same for both response curves. The other Hill parameters, y_∞, s and e are fitted for both compounds individually. In case the asymptote parameter y_∞ is below zero for any of the two Hill curves, the conditional response of that compound is refitted to a two-parameter model with y_∞ set to zero and y_0 kept from the fitting of both compounds together. This is the case for 43 records of the Mathews Griner dataset and 125 records of the Cokol dataset. We exclude records for which any of the Hill curve parameters slope or EC50 are negative ($s < 0, e < 0$). This is the case for 187 records for the Mathews Griner dataset (133 records with negative slope s, 88 records with negative EC50 value e, out of which there are 34 records with negative slope and negative EC50 value), which is roughly 40% of all records. More details follow below.

The $f_{\mathrm{GI}}(x_1, x_2)$ model is an implicit model for the response y. Therefore, a root finder is used to find a response $\hat{y}^{(i)}$ given concentrations $\left(x_1^{(i)}, x_2^{(i)}\right)$ and parameters describing the Hill curves of the conditional responses, $\Theta = \{y_0, y_{\infty,j}, e_j, s_j\}$, We used the standard implementation of a root finder in the R stats package, uniroot() [R Core Team, 2016], which is based on the Brent-Dekker-van Wijngaarden algorithm [Press et al., 2007, Chapter 9]. As convergence criterion we used 1.22×10^{-4} within a maximum of 1000 iterations.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

B.1 Sensitivity of model performance to inter-quantile range

In a previous version of this article, we cleaned the data to three times the inter-quantile range instead of five. With this smaller inter-quantile range we removed in the Mathews Griner dataset in total 199 data points instead of 23, and in the Cokol dataset 623 instead of 129. The performance of the Mathews Griner dataset for the lack-of-fit method slightly decreased, whereas the overall performance for the Cokol dataset increased (see Appendix E, Fig. 8 and Fig. 9).

As a note, approximately the same number of records were excluded for the analysis due to two issues: i) negative slopes of at least one of the conditional dose response curves, or ii) the root-finder for the $f_{\mathrm{CI}}(x_1, x_2|\alpha)$ model not converging (no convergence after 1000 iterations). These issues are independent from data cleaning with three or five times the inter-quantile range (see Appendix D, Table 13 - Table 16).

B.2 Handling records with negative slope or EC50 values

Roughly 40% (187) of the records of the Mathews Griner dataset were excluded in the study because of a negative slope or EC50 parameter. This is due to a suboptimal choice of doses. We observed two types of sub-optimality: first, the maximal dose can be too small to induce a significant change in response. Due to the noise in measurements, negative slope and EC50 parameters are fitted. This is the case for 34 records. A second type of sub-optimality is observed when the maximal effect is already reached for the second dose (the first dose is always zero). This is the case for the remaining 153 records.

Although we could not fit two reasonable Hill curves to these records, we can still use both methods, lack-of-fit and parametric, to quantify synergy. They both only require two mathematically well-defined conditional response curves. Here, we define a conditional response for cases with negative slope or negative EC50 parameter according to Table 2.

s	e	$y(x)$
< 0	< 0	$y_0 \quad \forall \quad x$
< 0	> 0	y_∞ for $x \neq 0$
> 0	< 0	y_∞ for $x \neq 0$

Table 2: Response curves for cases where the Hill model fit leads to negative slope or negative EC50 values.

With the above definition for conditional response curves, we investigated the 187 previously excluded records in detail. We computed the lack-of-fit synergy values γ for those 187 records and for the entire dataset of 466 records. For two of these records, the $f_{\mathrm{GI}}(x_1, x_2)$ model did not converge. The Kendall rank correlation coefficient values are given in Appendix D, Table 17. The inclusion of those datasets results in lower Kendall rank correlation coefficients relative to the original analysis. The coefficients decrease by roughly 0.05 when averaged over all models.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

C ROC-analysis

In high-throughput synergy studies, one generally screens for promising candidates that exhibit a synergistic or antagonistic effect. Those promising candidates are then investigated in more detail with genetic assays and other techniques. To determine how well the underlying null reference models result in distinguishable synergy scores, we conduct an ROC analysis (receiver operating characteristic), comparing the estimated synergy scores with the class categorization that is given for both datasets. A standard ROC analysis applies to binary classification, where cases are compared to controls. In this study, we have three classes: synergistic, antagonistic and non-interactive. We therefore compare each class to the combination of the other two, e.g. synergistic as cases versus the antagonistic and non-interactive combined as control. Typically, in ROC analyses, the cases rank higher than the controls. When treating the class antagonistic as case compared to the control synergistic and non-interactive we change all signs of the synergy scores. Therefore, the ranking of synergy scores is reversed and antagonistic synergy scores rank higher. Problems arise when comparing non-interactive cases to the control synergistic and antagonistic as their values should lie between the two control classes. Therefore, the absolute value of the estimated synergy scores is taken, which allows a ranking where the synergy scores of the non-interactive records should rank lower than the other synergy scores. Additionally, we can again multiply all synergy scores with minus one to revert the order of scores such that the cases rank higher.

The AUC values (area under the curve) are reported in Table 5 - Table 8 in Appendix D. For completeness, and based on the critique of Saito and Rehmsmeier [2015] to use PRC-AUC (precision/recall area under the curve) values for imbalanced datasets, the PRC-AUC values are also computed and can be found in Table 9 - Table 12 in Appendix D.

Analogously to the previous section, we depict the AUC values for both datasets in scatter plots (Fig. 10) with AUC values based on the parametric approach depicted on the x-axis and those based on the lack-of-fit approach on the y-axis. The underlying null reference models are shown by color. The different comparisons, such as synergistic versus non-interactive and antagonistic, are depicted by shape of the plot symbol.

From Fig. 10, the dominance of the lack-of-fit approach over the parametric one is as apparent for the Mathews Griner dataset as from Fig. 2. With regard to the comparison of the different cases, visualized in shape, the AUC values from the comparison of the synergistic cases to the non-interactive and antagonistic controls, score the highest values around 0.9. The comparison of the non-interactive cases to the interactive ones score the lowest.

As the overall highest AUC scores result from the lack-of-fit method, we have a closer look at those for both datasets (Table 6 and Table 8 in Appendix D). For the antagonistic case, the values range around 0.80 for the Mathews Griner dataset and around 0.85 for the Cokol dataset. AUC values of the non-interactive case range around 0.75 for both datasets. The AUC values for the synergistic case for both datasets range around a value of 0.90 with one outlier of 0.75 for the Bliss Independence model on the Mathews Griner dataset.

Overall, the lack-of-fit outperforms the parametric method on the Mathews Griner dataset. For the lack-of-fit method, both the $f_{\text{large}\to\text{small}}(x_1, x_2)$ and Explicit Mean Equation perform best on the Mathews Griner dataset for syner-

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

gistic cases, and a clear dominance of $f_{\text{large}\to\text{small}}(x_1, x_2)$ over the Explicit Mean Equation for antagonistic and non-interactive cases. On the second dataset, the Cokol dataset, Explicit Mean Equation performs overall best for both methods.

We attribute the differences in performances of methods and models on the two datasets to the differences in the experimental design for these datasets. For the Cokol dataset, all compounds were applied up to their maximal effect dose. In the Mathews Griner dataset, all compounds were applied with the same fixed dose range.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

D Supplementary Tables

model	lack-of-fit	parametric
$f_{\text{GI}}(x_1, x_2)$	0.52	0.23
$f_{\text{large}\to\text{small}}(x_1, x_2)$	**0.54**	**0.32**
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.48	0.19
$f_{\text{mean}}(x_1, x_2)$	0.53	0.10
$f_{\text{geary}}(x_1, x_2)$	0.47	0.30
$f_{\text{bliss}}(x_1, x_2)$	0.36	0.22

Table 3: Kendall rank correlation coefficient of Mathews Griner data set.

model	lack-of-fit	parametric
$f_{\text{GI}}(x_1, x_2)$	0.62	0.12
$f_{\text{large}\to\text{small}}(x_1, x_2)$	0.61	0.50
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.50	0.57
$f_{\text{mean}}(x_1, x_2)$	**0.67**	**0.64**
$f_{\text{geary}}(x_1, x_2)$	0.56	0.58
$f_{\text{bliss}}(x_1, x_2)$	0.56	0.16

Table 4: Kendall rank correlation coefficient of Cokol data set.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.67	0.63	0.62
$f_{\text{large}\to\text{small}}(x_1, x_2)$	**0.73**	**0.68**	**0.72**
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.68	0.53	0.53
$f_{\text{mean}}(x_1, x_2)$	0.62	**0.68**	0.46
$f_{\text{geary}}(x_1, x_2)$	0.70	0.57	0.60
$f_{\text{bliss}}(x_1, x_2)$	0.66	0.45	0.62

Table 5: AUC analysis of parametric method applied to Mathews Griner dataset.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.88	0.77	0.81
$f_{\text{large}\to\text{small}}(x_1, x_2)$	**0.89**	**0.78**	**0.84**
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.86	0.68	0.78
$f_{\text{mean}}(x_1, x_2)$	**0.89**	0.75	0.82
$f_{\text{geary}}(x_1, x_2)$	0.85	0.69	0.78
$f_{\text{bliss}}(x_1, x_2)$	0.75	0.60	0.76

Table 6: AUC analysis on lack-of-fit method applied to Mathews Griner dataset.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.62	0.62	0.55
$f_{\text{large}\to\text{small}}(x_1, x_2)$	0.80	0.64	0.84
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.86	0.64	0.89
$f_{\text{mean}}(x_1, x_2)$	**0.89**	**0.74**	**0.93**
$f_{\text{geary}}(x_1, x_2)$	0.80	**0.74**	0.89
$f_{\text{bliss}}(x_1, x_2)$	0.62	0.55	0.59

Table 7: AUC analysis on parametric method applied to Cokol dataset.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.93	0.78	0.88
$f_{\text{large}\to\text{small}}(x_1, x_2)$	0.94	0.63	0.86
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.83	0.66	0.83
$f_{\text{mean}}(x_1, x_2)$	**0.95**	**0.80**	**0.91**
$f_{\text{geary}}(x_1, x_2)$	0.88	0.71	0.86
$f_{\text{bliss}}(x_1, x_2)$	0.86	0.43	0.87

Table 8: AUC analysis on lack-of-fit method applied to Cokol dataset.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.39	0.69	**0.35**
$f_{\text{large}\to\text{small}}(x_1, x_2)$	**0.62**	0.73	0.32
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.52	0.58	0.13
$f_{\text{mean}}(x_1, x_2)$	0.56	**0.74**	0.18
$f_{\text{geary}}(x_1, x_2)$	0.61	0.61	0.15
$f_{\text{bliss}}(x_1, x_2)$	0.39	0.55	0.17

Table 9: PRC-AUC analysis on parametric method applied to Mathews Griner dataset.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.78	0.75	0.48
$f_{\text{large}\to\text{small}}(x_1, x_2)$	**0.80**	0.74	**0.55**
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.72	0.69	0.33
$f_{\text{mean}}(x_1, x_2)$	0.78	**0.76**	0.42
$f_{\text{geary}}(x_1, x_2)$	0.71	0.70	0.35
$f_{\text{bliss}}(x_1, x_2)$	0.52	0.64	0.39

Table 10: PRC-AUC analysis on lack-of-fit method applied to Mathews Griner dataset.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.30	0.57	0.51
$f_{\text{large}\to\text{small}}(x_1, x_2)$	0.56	0.56	0.65
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.62	0.52	0.83
$f_{\text{mean}}(x_1, x_2)$	**0.71**	0.55	**0.89**
$f_{\text{geary}}(x_1, x_2)$	0.60	**0.58**	0.82
$f_{\text{bliss}}(x_1, x_2)$	0.30	0.45	0.50

Table 11: PRC-AUC analysis on parametric method applied to Cokol dataset.

	synergistic	non-interactive	antagonistic
$f_{\text{GI}}(x_1, x_2)$	0.84	0.65	0.83
$f_{\text{large}\to\text{small}}(x_1, x_2)$	**0.87**	0.46	0.72
$f_{\text{small}\to\text{large}}(x_1, x_2)$	0.63	0.56	0.72
$f_{\text{mean}}(x_1, x_2)$	**0.87**	**0.66**	**0.86**
$f_{\text{geary}}(x_1, x_2)$	0.75	0.60	0.77
$f_{\text{bliss}}(x_1, x_2)$	0.76	0.35	0.70

Table 12: PRC-AUC analysis on lack-of-fit method applied to Cokol dataset.

	synergistic	antagonistic	non-interactive	total
parametric	19	15	48	82
lack-of-fit	34	59	93	186
both	19	16	49	84

Table 13: # of excluded records from parametric and lack-of-fit method applied to the Mathews Griner dataset with a threshold of three times the inter-quantile range.

	synergistic	antagonistic	non-interactive	total
parametric	21	16	49	86
lack-of-fit	36	58	91	185
both	21	16	49	86

Table 14: # of excluded records from the parametric and lack-of-fit method applied to the Mathews Griner dataset with cleaned data of a threshold of five times the inter-quantile range.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

	synergistic	antagonistic	non-interactive	total
parametric	6	4	6	16
lack-of-fit	7	5	7	19
both	6	4	6	16

Table 15: # of excluded records from parametric and lack-of-fit method applied to the Cokol dataset with cleaned data of a threshold of three times the inter-quantile range.

	synergistic	antagonistic	non-interactive	total
parametric	3	2	4	9
lack-of-fit	3	3	4	10
both	3	2	4	9

Table 16: # of excluded records from parametric and lack-of-fit method applied to Cokol dataset with cleaned data of a threshold of five times the inter-quantile range.

model	subset of 185 records	entire set of 464 records	original analysis on 279 records
$f_{\text{GI}}(x_1, x_2)$	0.37	**0.47**	0.52
$f_{\text{large}\to\text{small}}(x_1, x_2)$	0.34	0.43	**0.54**
$f_{\text{small}\to\text{large}}(x_1, x_2)$	**0.42**	0.46	0.48
$f_{\text{mean}}(x_1, x_2)$	0.37	0.43	0.53
$f_{\text{geary}}(x_1, x_2)$	0.35	0.39	0.47
$f_{\text{bliss}}(x_1, x_2)$	0.27	0.34	0.36

Table 17: Kendall rank correlation coefficients with recomputed conditional response curves according to Table 2 on 185 records with negative slope or EC50 value (left) and on entire dataset with 464 records (right).

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

E Supplementary Figures

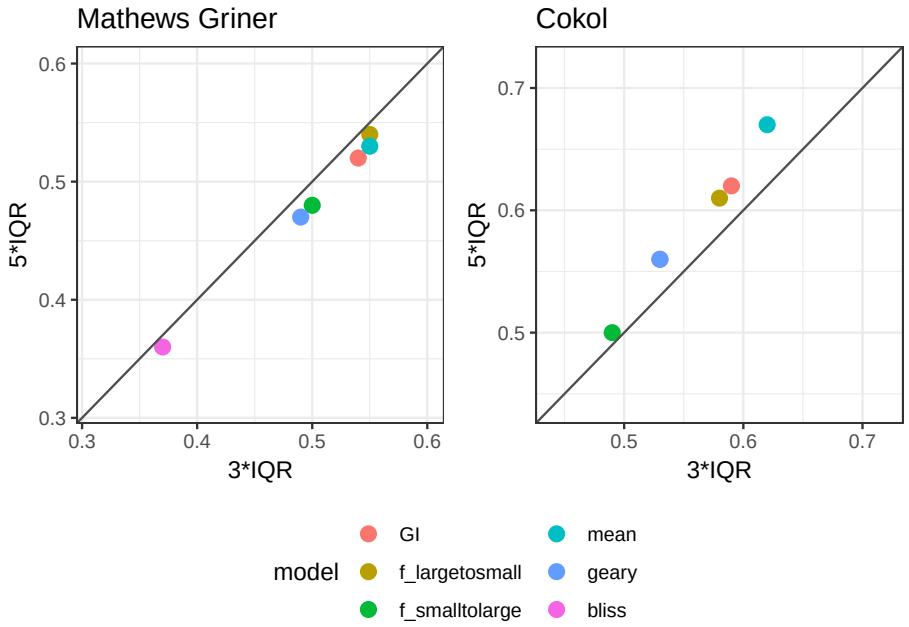

Figure 8: Scatter plot of Kendall rank correlation coefficient for both datasets, Mathews Griner (left) and Cokol (right), comparing the performance of the lack-of-fit method for different data-cleaning thresholds. The Kendall rank correlation coefficient values resulting from the cleaned data with three times the inter-quantile range are plotted on the x-axis and those from the data cleaned with a threshold of five times inter-quantile range are plotted on the y-axis. Each model is depicted in a different colour. To guide the eye, the diagonal is plotted.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

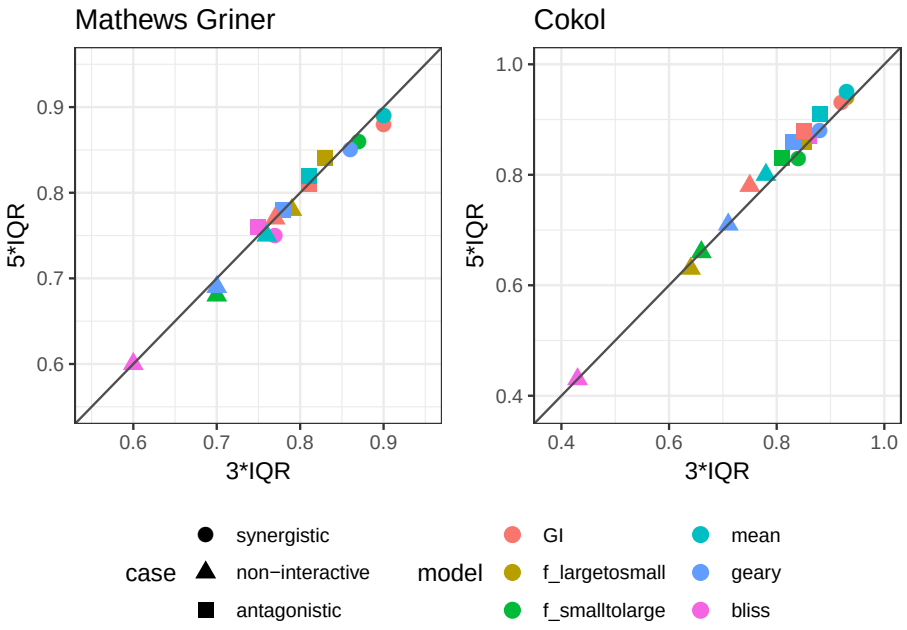

Figure 9: Scatter plot of ROC analysis for both datasets, Mathews Griner (left) and Cokol (right), comparing the performance of the lack-of-fit method for different data-cleaning thresholds. The AUC values resulting from the cleaned data with three times the inter-quantile range are plotted on the x-axis and those from the data cleaned with a five times inter-quantile range are plotted on the y-axis. AUC values from different models are shown in different colors. AUC values comparing the different categories are depicted in different shapes, where the naming of the shape represents the category that is compared to the remaining two. To guide the eye, the diagonal is plotted. The more a datapoint is above the diagonal, the better the performance of the data cleaned with a threshold of five times the inter-quantile range, and vice versa.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

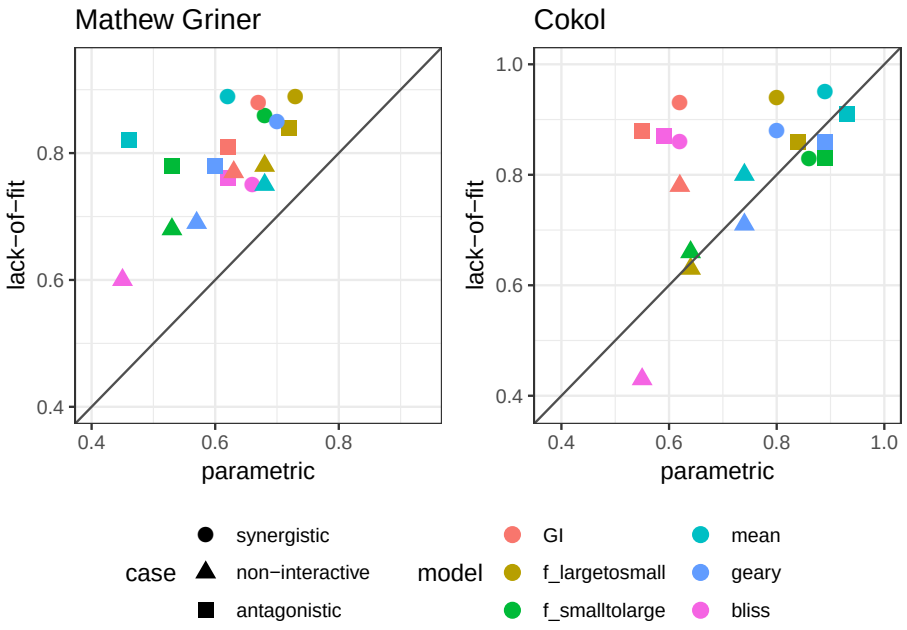

Figure 10: Scatter plot of ROC values for both datasets, Mathews Griner (left) and Cokol (right). The ROC values resulting from the parametric approach are plotted on the x-axis and those from the lack-of-fit approache are plotted on the y-axis. Each model is depicted in a different color. The three different comparisons, of one case versus the remaining two, are depicted in different shapes. To guide the eye, the diagonal is plotted. If a data point is above the diagonal, the ROC value from the lack-of-fit method is higher than that from the parametric method, and vice versa. Except for the non-interactive comparison of the Bliss Independence model, the synergy scores γ from the lack-of-fit method always result in higher ROC values than those computed based on the synergy scores α from the parametric method.

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

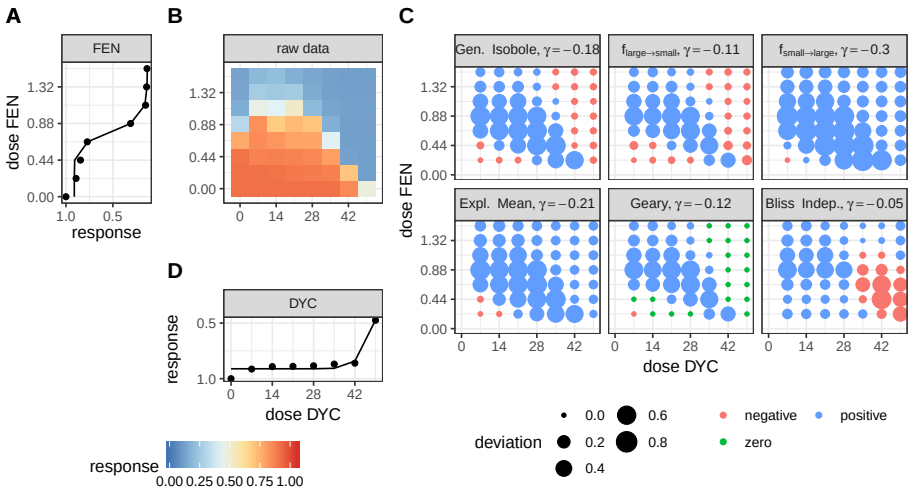

Figure 11: Description of the analysis steps of the lack-of-fit method for the compound pair FEN and DYC from the Cokol dataset. This compound pair is categorized as antagonistic according to [Cokol et al., 2011]. The raw response data of the record is depicted in (B). The response data normalized by the read at zero dose concentration (lower left). In (B) the degree of relative cell growth is colored from high to low values in red to blue.
Step 1: compute Hill curves for conditional responses: Based on the raw reads of the single dose responses (lower and left outer edges) fit a Hill curve to the conditional responses. The fitted Hill curves shown in (A) and (D) with original raw data shown as points.
Step 2: compute expected non-interactive response for all six models: not shown.
Step 3: compute difference between measured data (C) and expected data from all six null reference models: shown in (C). The direction of difference is shown by color (red for negative and blue for positive, green for zero). The larger the degree of difference, the larger the bullet, and vice versa.
Step 4: compute integral γ over the differences: Over all those bullets, we then compute the integral, which gives the synergy score γ. For every model, the synergy score γ is depicted in the title of each matrix in (C).

bioRxiv preprint doi: https://doi.org/10.1101/480608; this version posted July 31, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

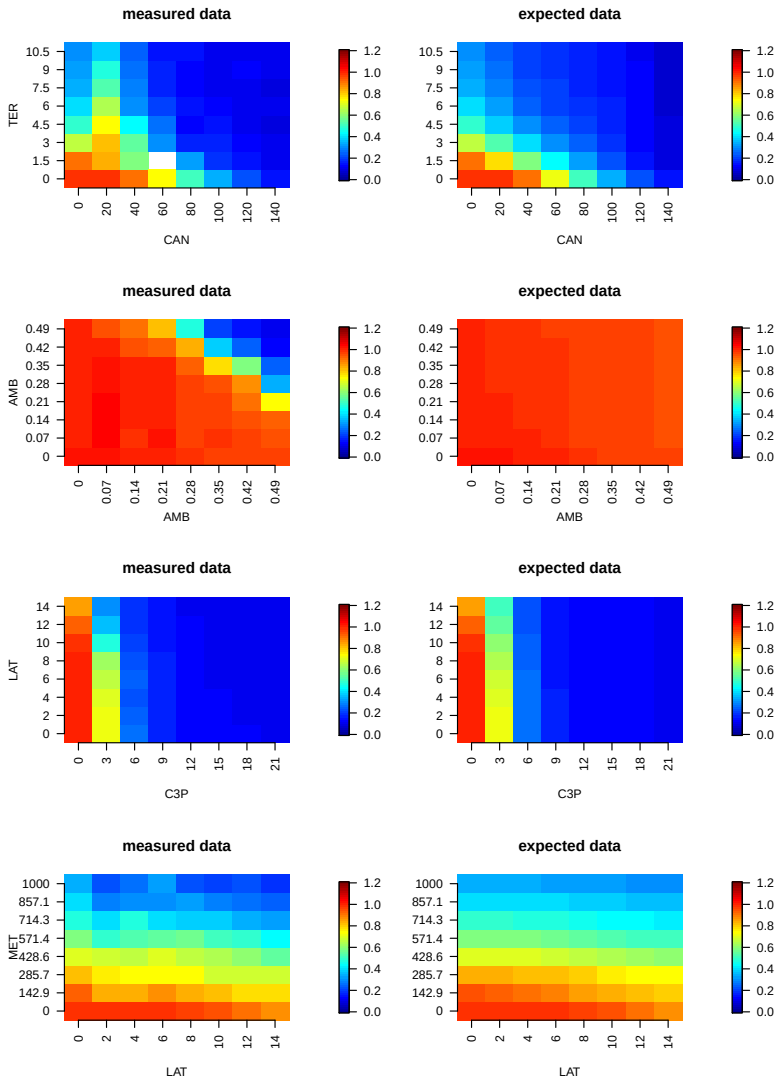

Figure 12: Raw responses (left) and expected responses from the Explicit Mean Equation model (right) of the four records from the Cokol dataset, for which the General Isobole Equation and Explicit Mean Equation gave synergy scores of opposite sign to the orignal categorization. More details on some parameters of the Hill curves can be found in Fig. 6.