Multilevel models improve precision and speed of IC_{50} estimates

Aim: Experimental variation in dose–response data of drugs tested on cell lines result in inaccuracies in the estimate of a key drug sensitivity characteristic: the IC_{50} . We aim to improve the precision of the half-limiting dose (IC_{50}) estimates by simultaneously employing all dose–responses across all cell lines and drugs, rather than using a single drug–cell line response. **Materials & methods:** We propose a multilevel mixed effects model that takes advantage of all available dose–response data. **Results:** The new estimates are highly concordant with the currently used Bayesian model when the data are well behaved. Otherwise, the multilevel model is clearly superior. **Conclusion:** The multilevel model yields a significant reduction of extreme IC_{50} estimates, an increase in precision and it runs orders of magnitude faster.

First draft submitted: 6 October 2015; Accepted for publication: 2 February 2016; Published online: 16 May 2016

Keywords: dose–response • IC_{50} • mixed effects • nonlinear

Personalization of cancer treatment is a major research effort aimed at identifying which patient is most likely to respond to a given treatment [1,2]. The mainstream approach to achieve personalized therapy is to stratify patients based on their tumor's molecular characteristics and administer drugs that exploit these. The current challenge is therefore to improve the understanding of how molecular characteristics of a tumor modulate drug response. Screening a large number of genomically well characterized cell lines against a large number of cancer drugs provides a unique opportunity to link given molecular features, such as gene mutations, to the effectiveness of a specific drug [3,4]. This knowledge allows selecting the optimal therapy for selected subpopulations of tumors, potentially avoiding ineffective treatments. Furthermore, associations between genetic features and drug response can provide valuable new information about the mechanisms of drug action.

Testing of drug sensitivity has been a routine procedure in research laboratories for

decades. It typically entails exposing cells to different concentrations of a drug and scoring the effects on viability relative to untreated controls. This relative viability is generally assumed to follow a sigmoid response as a function of the logarithm of the drug concentration. Fitting of the function underlying the dose–response can be then used to estimate the half-inhibitory concentration (IC_{50}), a commonly employed sensitivity statistic.

While the scale of screening efforts was previously limited by the large amount of manual work involved, robotics allows testing of an ever increasing number of compounds with unprecedented reproducibility. However, the methods for sensitivity estimation from dose–response data have not scaled with the technological progress. The unprecedented size of the data no longer allow for every individual fit to be inspected by eye to identify possible anomalies in the data points or the fitting results, yet existing methods have limited diagnostic quality in

Daniel J Vis^{*,1,2}, Lorenzo Bombardelli³, Howard Lightfoot⁴, Francesco Iorio⁵, Mathew J Garnett^{*,4} & Lodewyk FA Wessels^{*,1,2,6}

¹Center for Personalized Cancer Treatment (CPCT), Utrecht/Amsterdam, The Netherlands

²Department of Molecular Carcinogenesis, The Netherlands Cancer Institute, Amsterdam, The Netherlands

³Department of Molecular Genetics, The Netherlands Cancer Institute, Amsterdam, The Netherlands

⁴Wellcome Trust Sanger Institute, Cambridge, United Kingdom

⁵European Molecular Biology Laboratory – European Bioinformatics Institute, Cambridge, United Kingdom

⁶Technical University Delft, Delft, The Netherlands

*Author for correspondence: d.vis@nki.nl

[†]Authors contributed equally

this respect. On the positive side, the increased data size allows the borrowing of information across screens in order to improve the fit of sensitivity estimates. Therefore, we argue that results from large-scale drug screens should be analyzed in their entirety to provide the best possible sensitivity estimates. The approach we propose here not only increases the accuracy of estimates of the IC_{50} s but also speeds up the fitting procedure significantly, and can automatically detect poor model fits possibly caused by failed experiments.

The problem

An important issue related to curve fitting is illustrated in Figure 1. The two leftmost dose–response curves (III, I) can be fitted with ease but there is just a single observation in the rightmost dose–response (II) that shows a reduction in viability. Furthermore, this observation is not in the linear phase of the response, that is, the region that spans the 75 to 25% viability (approximately). As a result, many different sigmoids may fit the dose–response data almost equally well, and slight variations due to measurement noise, thereby substantially affecting the estimated IC_{50} value. In these

cases, one typically draws on additional information to obtain the best estimate. Garnett *et al.* [3] employ a Bayesian approach (from here on referred to as the Bayesian model) to each dose response curve individually. Priors constrained the range of the IC_{50} relative to the testing concentration and the shape parameters. These priors require tuning before the method can be applied. In contrast, the approach proposed here capitalizes on the other dose–response observations of the cell line to improve on the estimate for a specific compound. The model assumes that the slope of the dose–response curve is a cell line property and does not change with the specific compound. Each compound only influences the position of the center of the dose–response curve. This assumption adequately describes the data but it is an obvious simplification. Still, there are three advantages of this modeling approach: it extracts the most out of the available observations by borrowing information across cell lines and drugs; it alleviates the problem of extreme estimates caused by local minima commonly seen in individual dose–response fitting; it allows for detecting failed experiments, and it offers a significant speed advantage over

Figure 1. Sigmoid dose–response relation for a mock compound. Black line (I) shows the dose response function with a linear phase in the center of the sigmoid. Each series spans 256-fold concentration range, and each unit translates to a twofold difference, meaning that 7 to 9 translates to a fourfold difference. Changes to the center of the function define the change in IC_{50} , which is the concentration at which the relative viability is 0.5 (horizontal line). Examples of a less potent drug (II) and a more potent drug (III) are shown with dashed lines. The inset (AUC) illustrates the area under the curve in the screening range, and is defined as $A/(A + B)$. AUC: Area under the curve.

Bayesian approaches that rely on sampling procedures such as Markov-chain Monte Carlo.

Methods

The data

Drug sensitivity data comprised 79,903 dose–response series distributed over 707 cell lines and 145 drugs were downloaded from the Genomics of Drug Sensitivity in Cancer Website version 5 (October 2013, [5]) [6]. The data generation is described in Garnett *et al.* [3]. Briefly, compounds were provided by academic collaborators or were sourced from commercial vendors. The range of concentrations selected for each compound was based on *in vitro* data of concentrations inhibiting relevant kinase activity and cell viability, as well as clinical data indicating peak and trough plasma concentrations in human subjects. Each dilution series spanned a 256-fold concentration range, as either a 9-point two-fold or a 5-point fourfold dilution series plated by liquid handling robotics. The quantification of the fluorescent signal intensity for the viability assessment was performed using a fluorescent plate reader at the appropriate excitation wavelengths. Each plate contained cell-line specific positive (no viable cells) and negative (no treatment) controls to which the treated cells were compared. Further details on the data generation are described in [Supplementary Material Section 2.1](#) and Garnett *et al.* [3].

Additionally, we employ a small bespoke screen published by Calbo *et al.* [7], from here on the Calbo small-cell data, to illustrate the application of our approach to a drug sensitivity screen of a size more frequently encountered in research labs. Six well-characterized mouse small-cell lung cancer cell lines were tested for their sensitivity to four anticancer drugs. The experiments were performed in triplicate, sampling the concentration range at nine values, with a fourfold dilution between consecutive concentrations.

Relative viability

The relative viability was assessed by scaling the signal of the treated wells to a set of untreated but seeded wells (100% viability, negative control), and a set of wells with no cells seeded (0% viability, positive control). The mean fluorescent intensity for both types of controls are computed and used for scaling the intensities of the treated cells. This gives the viability estimate, relative to the controls. [Equation 1](#) summarizes this step, in which y_{ij} is the observed relative viability for cell line i , and drug j at concentration x , p_i and n_i are the (average) positive and negative control intensities and $\tilde{y}_{ij}$ is the observed absolute intensity.

$$y_{ijx} = \frac{\tilde{y}_{ijx} - n_i}{p_i - n_i}$$

Equation 1

The viability values are clipped at the 0 and 1 bounds. The above preprocessing is applied to all cell lines and drugs, comprising a data set of over 79,903 combinations.

The model

It is assumed that the cell viability is a function of the drug concentration and that this function is adequately described by a two-parameter sigmoidal function. In the model, one parameter determines the IC₅₀ value (position) relative to the maximum concentration used, and the other parameter determines the slope. Furthermore, it is assumed that the viability in the absence of the drug is 100%, and that, with a sufficiently high dose, 0% viability can eventually be reached. Another simplifying assumption is that the viability cannot be higher than 100% or lower than 0%.

The sigmoid dose–response model describes the viability with two free parameters (slope, position) as a function of the dilution step-range (x), see [Equation 2](#).

$$f(b_{\text{pos}i}, b_{\text{shape}i}, x) = \frac{1}{1 + e^{-\frac{x - b_{\text{pos}}}{b_{\text{shape}}}}}$$

Equation 2

The IC₅₀ is the x dilution step for which $f(b_{\text{pos}i}, b_{\text{shape}i}, x) = 0.5$. Both the parameters b_{pos} and b_{shape} are allowed to vary across the cell lines, while only b_{pos} varies across cell line-drug pairs by means of a drug specific effect, see [Equations 3 & 4](#). The β 's in [Equations 3 & 4](#) are the population estimates for the position and shape parameter. The b_{1i} and b_{2i} values represent the cell line level and describe the effect in the position and slope parameter for cell line i . The full covariance structure of the effect and slope parameters is estimated for the cell line level (see [Equation 5](#)). Finally, the b_{1ij} value represents the effect of the cell line drug combination for drug j in cell line i . Given the cell line specific effects, all drug responses are assumed to be generated by the same distribution, where the drug effect is nested in the cell line effect. The following set of equations defines the model:

$$b_{\text{pos}} = \beta_1 + b_{1i} + b_{1ij}$$

Equation 3

$$b_{\text{shape}} = \beta_2 + b_{2i}$$

Equation 4

$$\begin{bmatrix} b_{1i} \\ b_{2i} \end{bmatrix} \sim \mathcal{N}(0, \Psi)$$

Equation 5

$$b_{1ij} \sim \mathcal{N}(0, \sigma^2)$$

Equation 6

in which x is the dilution step, b_{pos} is the IC_{50} and b_{shape} is the slope parameter. Throughout, by convention, the response is modeled such that the maximum concentration is taken as $x = 9$. A unit reduction the x -axis corresponds to a twofold dilution.

Power is gained by formulating the model as a multilevel model with random effects. Such a model allows the inference of the shape of the cell line response across all drugs, which significantly improves parameter inference.

Algorithms

The nonlinear mixed effect model as described in Equations 2–6 was fitted using nlme [8] in R [9]. The approach can be used for sigmoid response functions for which multiple samples and treatments are available and outputs the position (IC_{50}) and slope estimates for every cell line drug combination.

The memory and CPU requirements are limited; an Intel i7 3xxx processor at 2.2 GHz with 8 GB of RAM is able to perform the described analysis in about 2 h. The use of optimized basic linear algebra subroutines (BLAS) is recommended. The results presented here employ the OpenBLAS (version 0.2.8, based on GotoBLAS2 1.13 BSD), using R 3.1. The method and its convenience functions have been packaged and are available to the general public.

Results

The fit

The model describes the population of dose–response curves. Each individual cell-line drug pair is described as a deviation of the population estimate. The population dose response curve is shown in Figure 2A and is defined by the population parameters β_1 and β_2 from Equations 3 & 4. To obtain the cell line specific curve, the cell line specific position and slope deviation ($\beta_1 + b_{1i}$, $\beta_2 + b_{2i}$) are added to these population values. The scatterplot of these cell-line specific parameter values is shown in Figure 2B. These parameters are correlated and this means that the more sensitive lines tend to have a steeper slope. The cell line specific values for the HEL and LN-405 cell lines are indicated in red and black dots, respectively, in Figure 2B. Based on these values we can construct the cell line specific

dose response curves (characteristic curves) depicted in Figure 2C. Finally, inclusion of the drug-effect (b_{1ij}) shifts these characteristic dose response curves to account for the drug effects, see Figure 2D. The parameters carry the following biological interpretation; the slope of the response to a dosage escalation is dominated by a cell line specific effect, while the position of the IC_{50} is a drug-dependent effect, see also Supplementary Figure 2. Cell lines can be inherently more sensitive or resistant, relative to the population and this behavior is correlated with the slope but the drug effect is larger (estimated variance for the position parameter on the cell line level is 1.10, on the drug within cell line it is 8.76).

The model fit to the data is illustrated with the same lines on three different compounds, a PI3K (class 1) + mTOR1/2 inhibitor (BEZ235), an mTOR1/2 inhibitor (AZD8055) and the DNA-crosslinker cisplatin. Figure 2E & F show the fitted curves for HEL and LN-405, respectively. The underlying data points are shown as dots connected by dashed lines. The fitted curve is a good approximation of these data, with a root mean square error (RMSE) <0.15.

The two cell lines used in this example differ in their (estimated) intrinsic sensitivity, as shown in the scatter plot of the cell line variation (Figure 2B). This sensitivity is reflected in the dose–response curves that show lower viability (response activity) across the screening range for HEL than for LN-405 (Figure 2E & F). Cisplatin at the same dose affects HEL much more than LN-405, which is nearly resistant.

Interestingly, we find that the slope parameter is not uniformly distributed across the various tissue types represented in the cell lines panel. In particular, we find that in this screen the lymphomas are generally more sensitive and with a steep slope. Renal, lung and thyroid derived lines, on the other hand, tend to be more resistant and with a shallower slope, see also Supplementary Table 1.

Activity & model error

A widely used summary statistic of the dose–response curve is the area under the (dose–response) curve (AUC); small AUC values indicate strong response, whereas large values indicate limited or no response. The model error can be estimated by the root mean squared error (RMSE) between the area under the dose–response measurements and the area under the fitted curve. This is indicative of how far the model approximation deviates from the measured observations. In Figure 3, we further illustrate different response types based on the AUC. The light blue line/dots represent a dose–response with a large effect on cell number (low AUC), while the dark blue

Figure 2. Graphical presentation of the model parameters. (A) The population dose–response estimated by the model from the data. (B) Cell lines differ in their position (x-axis) and slope (y-axis) parameter. Cell lines that are more resistant tend to have a shallow slope. (C) Two characteristic dose–response curves for two cell lines (see [B]) are plotted, with the population response in gray (A). (D) A drug, cisplatin, illustrates how the characteristic response curve shifts (dashed lines) depending on its sensitivity, the black line is quite sensitive. (E & F) show real fits and the underlying data for the same cell lines for three drugs.

Figure 3. Illustration of responses for six randomly selected cell lines. For each line a response with good activity (light blue), limited to no activity (dark blue) is shown. The purple response is unexpected and is detected by the model error statistic root mean squared error.

line/dots show limited to no activity (high AUC). Both are fitted accurately by a sigmoid, as reflected in a low RMSE. The purple line/dots show a dose–response that has a high model error (high RMSE).

The RMSE is a very useful criterion to identify dose response curves that have poor model fits. An empirical evaluation of the data revealed that an RMSE of 0.3 or higher is associated with poor quality data due to excessive technical noise, or dose–responses that do not conform to the model's assumptions and are therefore poorly fitted.

When all cell line screens in combination with a given drug are associated with a consistently high model error it can indicate an inadequate model. We computed the average RMSE for each of the drugs in the screen and find that the top five drugs with the largest average residual are: rapamycin, shikonin, temsirolimus and the two experimental cancer drug candidates OSU-03012, and AUY922. Rapamycin and temsirolimus (a rapamycin derivative) both target mTORC1 and have previously been reported to have unusually shallow dose–response curves [10].

We find that many of the cell line responses for these two drugs are almost flat and show very little difference between the lowest and the highest drug concentration. See for example [Figure 3](#) where LN-405 on temsirolimus shows a flat dose–response curve when we consider the measurements data only. This behavior is not observed for mTORC1/2 targeting agents, suggesting that the mTORC1 targeting drugs rapamycin and temsirolimus may in fact result in nonsigmoidal dose–response curves. Alternatively, the metabolic viability assay utilized and the mTORC1 drugs may interact leading to dose–response artifacts. We find little evidence to support previous observations that drugs targeting the mTOR/AKT/PI3K pathway are generally associated with a flat response [10]. Indeed, drugs targeting both mTORC1 and mTORC2 appear to be well approximated by the model, as are AKT and PI3K targeting drugs. For shikonin, OSU-03012 and AUY922, the larger residuals can be explained in terms of a much lower average AUC and noisy observations, rather than a structural problem with the approximation of the dose response curve.

Concordance & precision

The IC₅₀s and AUCs in Garnett *et al.* [3] were estimated with the Bayesian model. Here, we report how the estimates of that model compare to those obtained with our method. We find that both the IC₅₀ and AUC correlate well. The median correlation across all 145 drugs is 0.91 (range: 0.27–0.98) for IC₅₀ values and 0.92 (range: 0.67–0.98) for AUC values. The estimates are highly correlated, in particular for dose–response

curves that show clear activity (low AUC). For those that are associated with modest activity (average AUC value is high), the Bayesian model tends to overestimate the IC₅₀ relative to the multilevel approach advocated here. Lastly, the estimates of the Bayesian model include many more extreme high IC₅₀ values (due to local minima) compared with the new approach.

To determine if the multilevel model improves the precision of IC₅₀ estimates we utilized a dataset consisting of 41 drugs screened across six cell lines with three technical replicates. Each technical replicate is considered to be a single screen in the multilevel model giving three IC₅₀ estimates for each cell-line drug combination. The estimated model parameter are tabulated in [Supplementary Table 2](#).

The precision of estimates can be summarized by the coefficient of variation (CV) also known as the residual standard deviation. The CV is defined by the standard deviation of the replicated concentrations divided by their mean, that is $c_v = \sigma/\mu$. The coefficient of variation is computed on the actual concentration, rather than the log thereof. This reflects the exponential nature of the concentration test range and preserves the non-negative drug concentration values. The coefficient of variation is computed for each cell line and drug combination across technical replicates. The precision of the system is illustrated by comparing the CV values between the Bayesian model and the multilevel model. [Figure 4](#) shows the results of the analysis for each cell line. For all cell lines, the mean CV is lower in the multilevel model compared with the Bayesian model ($p < 0.01$ in all cases based on the paired t-test).

Scale

Although the multilevel approach is suitable for large-scale pharmacological datasets, we wished to determine if it could be applied to small-scale experiments of the type more typically performed in laboratories. Many laboratories use commercial software, such as Graphpad Prism (GraphPad Software, CA, USA), that fits a predefined sigmoid curve to normalized measured dose–response values. We compared this widely used approach to our multilevel approach.

We found a high degree of concordance between the two methods for the Calbo small cell data ([Supplementary Figure 1](#)) and both led to the same conclusion regarding the experimental outcome. This dataset comprises six cell lines and four drugs. The correlation between the IC₅₀ estimates is very good ($r \approx 0.99$), except for one drug that did not show any activity in the tested cell lines, resulting in a lower correlation ($r = 0.71$). These results indicate that the multilevel model performed at least as well as the accepted standard models in smaller screens.

Figure 4. The coefficient of variation was used to investigate the accuracy of estimates in a replicate series. The CV statistics are shown for all the lines for both approached. The multilevel model performs significantly better in all cell lines. CV: Coefficient of variation.

Speed

The high computational speed is an advantage of our model. The Bayesian model estimates about ten cell line-drug combinations per hour on a typical workstation, requiring about 8300 CPU hours to infer all IC₅₀'s values in this version of the Sanger panel [3]. The multilevel approach requires approximately an hour for the same task, thereby improving the computational speed by three orders of magnitude.

Discussion

Here, we presented a multilevel model that estimates the IC₅₀ values for all cell lines and compounds of a drug screening simultaneously. This model outperforms the Bayesian model used in Garnett *et al.* [3]. More specifically, it outputs fewer extreme IC₅₀ values, more precise IC₅₀ estimates at a computational speed that is 1000-times faster.

Increased precision is achieved by exploiting a property of drug response screens, the same cell line is screened against multiple drugs, and the same drug is screened against multiple cell lines. Hence, there is an opportunity to share information across cell line – drug combinations in order to reduce variance originating from differences in experimental conditions. For example, in our large drug screen data, we observed that the slope of the dose–response curve is similar for a cell line screened against many drugs. Therefore, instead of re-estimating the cell line's slope for every drug, we explicitly add the assumption that this slope is the same for all drugs. This yields a considerable increase in sample size, and therefore allows for a more reliable estimate of the slope parameter.

It is conceivable that some drugs induce a dose–response that is intrinsically different from the sigmoid model assumed. This may be due to noisy observations or inadequate model assumptions. With our model, these can be readily identified using the RMSE for each dose response curve. The first should be randomly distributed across cell line and drug combinations, while inadequate model assumptions are believed to be associated with the drug. We observed a large model error for drugs targeting mTORC1 consistent with earlier reports that mTORC1-targeting drugs show an unusually flat dose–response curve [10]. Responses that cannot be adequately approximated using a sigmoid may need to be analyzed differently. Our method brings the advantage

that such cases can be identified systematically, even among thousands of tested cell line–drug combinations.

Conclusion

Our new multilevel approach for the analysis of drug response data capitalizes on all the information available to enable modeling of all dose response curves simultaneously. By combining information across the entire screen, more accurate estimates can be obtained without resorting to externally injected knowledge, for example, by using informative priors. This approach allows fitting of data with greater precision and speed than current approaches.

Future perspective

It is speculated that the field of large scale drug sensitivity testing will move toward larger screens and hence toward the multilevel mixed effect type models as presented here. These models offer an efficient route to formalize and integrate information from the entire set of dose–responses. This alleviates the computational burden, suppresses noise and simultaneously makes explicit where the modeling assumptions are no longer supported by the data.

Supplementary data

To view the supplementary data that accompany this paper, please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/pgs.16.15

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending, or royalties.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Executive summary

- Experimental error obscures dose–response curve.
- Set of dose–responses may be analyzed simultaneously.
- Multilevel mixed model takes advantage of all the data.
- Significantly improved precision, fewer extreme IC₅₀ estimates and 100+ fold faster runtime.

References

- 1 van der Greef J, Hankemeier T, McBurney RN. Metabolomics-based systems biology and personalized medicine: moving towards n = 1 clinical trials? *Pharmacogenomics* 7(7), 1087–1094 (2006).
- 2 Vis DJ, Westerhuis JA, Jacobs DM *et al.* Analyzing metabolomics-based challenge tests. *Metabolomics* 11, 50–63 (2015).
- 3 Garnett MJ, Edelman EJ, Heidorn SJ *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483(7391), 570–575 (2012).
- 4 Barretina J, Caponigro G, Stransky N *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483(7391), 603–607 (2012).
- 5 Wellcome Trust Sanger Institute, Genomics of Drug Sensitivity in Cancer Website. www.cancerRxgene.org
- 6 Yang W, Soares J, Greninger P *et al.* Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res.* 41(Database issue), D955–D961 (2013).
- 7 Calbo J, van Montfort E, Proost N *et al.* A functional role for tumor cell heterogeneity in a mouse model of small cell lung cancer. *Cancer Cell* 19(2), 244–256 (2011).
- 8 Lindstrom ML, Bates DM. Nonlinear mixed effects models for repeated measures data. *Biometrics* 46(3), 673–687 (1990).
- 9 R Core Team. *A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2016). www.R-project.org/
- 10 Fallahi-Sichani M, Honarnejad S, Heiser LM, Gray JW, Sorger PK. Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nat. Chem. Biol.* 9(11), 708–714 (2013).