

Beyond IC_{50} s: Towards Robust Statistical Methods for *in vitro* Association Studies

Andrew Beam¹ and Alison Motsinger-Reif^{1,2*}

¹Bioinformatics Research Center, North Carolina State University, Raleigh NC, USA

²Department of Statistics, North Carolina State University, Raleigh NC, USA

Abstract

Cell line cytotoxicity assays have become increasingly popular approaches for genetic and genomic studies of differential cytotoxic response. There are an increasing number of success stories, but relatively little evaluation of the statistical approaches used in such studies. In the vast majority of these studies, concentration response is summarized using curve-fitting approaches, and then summary measure(s) are used as the phenotype in subsequent genetic association studies. The curve is usually summarized by a single parameter such as the curve's inflection point (e.g. the EC/IC_{50}). Such modeling makes major assumptions and has statistical limitations that should be considered. In the current review, we discuss the limitations of the EC/IC_{50} as a phenotype in association studies, and highlight some potential limitations with a simulation experiment. Finally, we discuss some alternative analysis approaches that have been shown to be more robust.

Discussion

Association mapping (with either DNA level variation or gene expression data) in pharmacogenomics has been impeded by the reliance on clinical trials for samples.

Genetic studies nested within clinical trials face the limited ability to enroll enough human subjects, ethical constraints, and the presence of uncontrolled confounders all of which may limit the capability to identify loci involved in drug-response [1]. To address these limitations, *in vitro* association studies have been proposed as an alternative to human-based studies because they address many of these concerns for certain types of drugs [1,2]. Cell-based studies offer extremely large sample sizes and do not require approval from regulatory agencies, resulting improved statistical power while decreasing both time and cost needed to conduct a study. Consequently they allow for the rapid study of drug response in a highly human relevant system for a fraction of the expense of traditional methods. Moreover, these assays can be made tissue or disease specific by using cultures of the relevant cell type, further increasing *in vivo* relevancy. A more detailed discussion of the advantages and limitations of such *in vitro* assays has previously been reviewed [1,2]. An increasing number of success stories for such experiments are emerging [3-9], but the statistical methodologies applied in such experiments have not been examined in detail.

Cell-based studies allow for the examination of drug response at greater resolutions by measuring cellular response across a spectrum of concentrations rather than limited set of concentrations afforded by traditional studies. These types of *dose-response* or *concentration-response* studies measure some indication of cellular health or response such as total ATP, cell viability/morphology, or transcript expression levels as a function of increasing drug concentration. These data points are then fit to a statistical model, usually some form of a 4-parameter logistic curve (sometimes referred to as the *hill equation*), to produce a dose-response curve. Figure 1 shows an example of an annotated concentration-response curve. The curve is usually summarized by a single parameter such as the curve's inflection point (e.g. the EC/IC_{50}) [10] or the slope of the curve (called the *hill-slope*) [11]. Perhaps the most widely used summary in pharmacogenomics cell line experiments is the IC_{50} , which represents the concentration where the response achieves 50% of maximal activity [3-9]. This notion of IC_{50} can be generalized further such that IC_x is the concentration at which

the response is X% between minimal and maximal activity. IC_{50} s (and their IC_x cousins) have been widely used in areas such as toxicology, pharmacology/pharmacogenomics, and industrial drug development [10]. Its popularity derives from the fact that it is a concise and interpretable summary of a drug's activity, which conveys an indication of the drug's *potency*. In association studies, this value is treated as a quantitative trait and standard QTL methods are then applied to link genotype to this derived phenotype [12].

However, traditional analytic and statistical methods are often ill-equipped to analyze this type of data and subsequent inference based on the IC_{50} poses many challenges. First, the appropriateness of the hill-slope model from which the IC_{50} is derived is often unchecked, which may have large implications on the resulting conclusions [12]. This model is based on ligand-macromolecule binding dynamics [13], which may be an appropriate model in some instances, but inappropriate in others. Assuming this model is a correct description of the underlying biology, accurate calculation of the IC_{50} may still prove problematic. Estimating the IC_{50} using this model is highly sensitive to observing the full dose-response curve in the tested concentration range. If either the minimum or maximum asymptote of this curve is not observed it can have a very large impact on the estimated IC_{50} which will have a correspondingly large impact on the biological conclusions. Due to the non-linearity of this model even the "well-behaved" responses may result in unstable IC_{50} estimates. Two analysts may reach different IC_{50} values because they used different software packages or because they used the same software with different configuration settings. Such differences may cause the software to fail to produce a solution at all or

*Corresponding author: Alison A. Motsinger-Reif, Bioinformatics Research Center, 1 Lampe Drive, CB 7566, Ricks Hall, Raleigh, NC 27695, USA, Ph: 919-515-3574; E-mail: motsinger@stat.ncsu.edu

Received November 14, 2013; Accepted December 02, 2013; Published December 10, 2013

Citation: Beam A, Motsinger-Reif A (2013) Beyond IC_{50} s: Towards Robust Statistical Methods for *in vitro* Association Studies. J Pharmacogenomics Pharmacoproteomics 5: 120. doi:10.4172/2153-0645.1000120

Copyright: © 2013 Beam A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

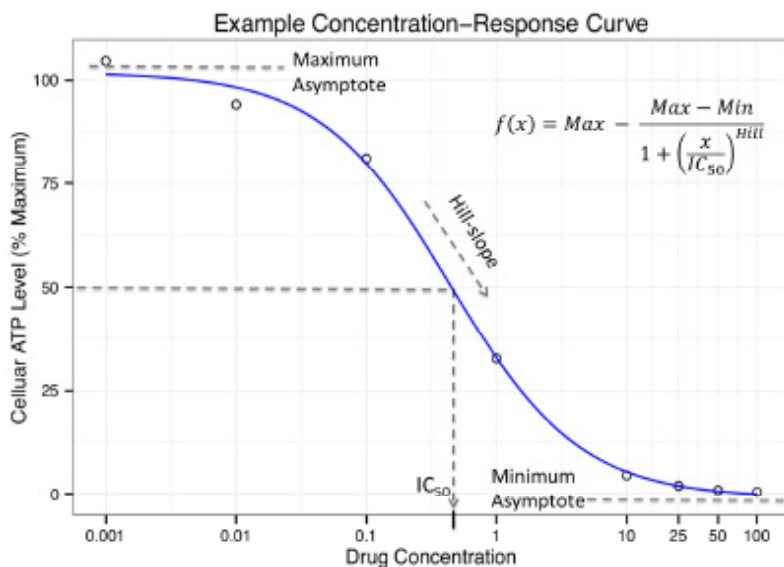


Figure 1: An example concentration-response curve with the 4 parameters (Maximum, Minimum, IC_{50} , and hill-slope) of the hill-slope model labeled. The equation is displayed in the upper-right corner.

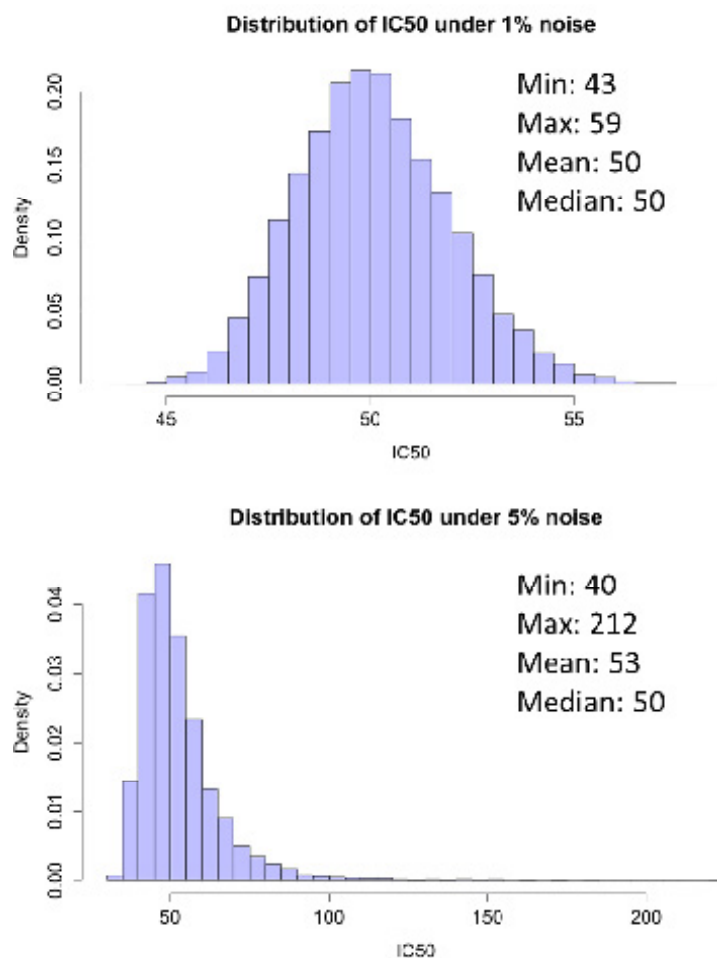


Figure 2: Distribution of the IC_{50} under two levels of noise. Note the wide range of estimated IC_{50} s, especially under the slightly higher 5% noise model with many estimates being 2-3x larger the true value of 50 μ M. Note also that the second histogram is no longer symmetric, implying that these IC_{50} s are not normally distributed.

may produce very different IC₅₀ estimates, with no clear procedure for determining the correct value [14].

For example, assume a study measures total ATP across 8 concentrations using five technical replicates. Some summary statistic of these replicates, such as the mean or median ATP level at each concentration may then be fit to the hill-slope model to obtain an IC₅₀ value. The amount of uncertainty introduced from sampling the response at each concentration can have a considerable impact on the estimated IC₅₀. To highlight this issue and give some sense of how it can affect the IC₅₀, we performed a small simulation experiment. For a concentration set of {0.0001,0.001,0.01,0.1,1,10,50,100} μM, we simulated 5 technical replicates for 2 levels of noise. Each technical replicate was generated from the hill-slope model plus a small amount of random noise with a true IC₅₀ value of 50 μM. The noise was a random value of +/- X% of the true response value, where the values for X we tested were 1% and 5%. This is a 'heteroskedastic' noise model and is consistent with our experience the amount of variation in a response is proportional the size of the response itself. We then took mean of the 5 technical replicates and fit a curve using the *nls()* function in the R statistical language [15] to this mean response and recorded the estimated IC₅₀. Note that we supplied the algorithm with true parameter values as starting values so as to minimize the amount of IC₅₀ variation coming from the fitting process. We repeated this process 10,000 times for both levels of noise. Figure 2 shows histograms for the 1% and 5% noise level models. Even in the presence of a small amount of noise (1%), the IC₅₀ estimates span a range of 40-70 μM and inspection of the estimated confidence interval for one such response yields a similar estimated range of variability. The situation for higher noise model is much worse with the IC₅₀ estimate ranging from 40-212 μM, which was outside of the tested concentration range. This amount of uncertainty results in an IC₅₀ measure that is not very useful in practice because it is statistically indistinguishable from a potentially wide range of other IC₅₀ values. In the context of association studies, this could be of great harm. Imagine that there are two populations where one population has an *estimated* IC₅₀ that is 2-3x that of the reference population, indicating that this population may be highly tolerant to the drug under study. It would be of great interest to locate any genetic loci that may be involved in this process. However, as the Figure 2 implies, these two populations may in fact have the same tolerance for the drug, but the noise introduced through sampling and estimating the IC₅₀ has obfuscated our ability to see this, resulting in wasted time and effort looking for the underlying causes of a phantom difference.

This leads to yet another issue with IC₅₀ based inference, namely that once all of the proper variation is accounted for, IC₅₀s may show little meaningful variation in the statistical sense. This may result in two compounds, which by other measures would be considered to have different activity, to fail to be declared distinct, because their IC₅₀s are not statistically separable. Statistical models are built to explain variation, but in the absence of meaningful variation, they will be unable to detect any genetic signal that may be present. This will be of increasing importance if pharmacogenomics, and genomics more generally, is to unravel complex traits that do not have large, single gene effects.

A somewhat larger point worthy of consideration is just how relevant an IC₅₀, even one estimated with high precision, is to the underlying scientific question. Statistical methods are only valid to the extent to which they mapback to the research question being asked. Even in the absence of all the issues discussed so far with IC₅₀ based inference, it may be that a "true" statistically significant difference is not

very meaningful from a biological perspective. Why might we assume *a priori* that *this* parameter from *this* model is the best representation of a compound's activity? In this sense, it is not clear that IC₅₀s are always a relevant measure or summary of a compound's activity, if potency is not a meaningful proxy for the latent biological difference. If the IC₅₀ is a poor proxy, then methods that take the full dose-response into consideration should be considered.

With both the promise of *in vitro* studies and the analytic challenges they present in mind, we hope to draw attention to some of the issues that must be addressed to maximize the utility of these types of assays. There are alternatives to the IC₅₀ based significance testing approach that have been and continue to be developed. The area under the curve (AUC) statistic computes the area between the dose-response curve and the x-axis and is a global measure of compound's activity [12]. This type of summary is potentially more robust than an interpolated parameter such as an IC₅₀. Determination of statistical difference between two compound's AUC relies on a permutation testing based procedure and may be very computationally expensive for large datasets. However, since permutation testing can be readily parallelized, the availability of computing clusters can reduce the time needed for this type of analysis. Multivariate ANOVA Genome-Wide Association Software (MAGWAS) [16] was shown to be a very attractive approach with many desirable properties including high statistical power and computational efficiency. However, MAGWAS is sensitive to changes that occur only at one concentration, which may not be desirable in some instances. Both AUC and MAGWAS incorporate the full dose response curve into the association tested.

Each of these concerns is only exaggerated by the increasingly high throughput nature of these experiments. As robotics has enabled rapid, high-throughput phenotyping for such experiments, investigators are now able to readily assay dozens or even hundreds of chemicals across hundreds of cell lines for dose response [17]. This makes it less likely that all assumptions are met or checked across such large numbers of results. This magnifies the importance of considered statistical approaches that minimize the impact of violations from these assumptions.

It is our hope that this discussion can help further the continued consideration on best practices for *in vitro* association studies. While we acknowledge that that IC₅₀s can be of great utility when used properly and in the correct context, we hope to raise awareness of potential issues with these approaches and highlight alternatives that could further our understanding of gene based drug response.

Acknowledgment

This work was supported by NIH NCI RO1CA161608 from the National Cancer Institute.

References

1. Welsh M, Mangravite L, Medina MW, Tantisira K, Zhang W, et al. (2009) Pharmacogenomic discovery using cell-based models. *Pharmacol Rev* 61: 413-429.
2. Wheeler HE, Dolan ME (2012) Lymphoblastoid cell lines in pharmacogenomic discovery and clinical translation. *Pharmacogenomics* 13: 55-70.
3. Watson VG, Motsinger-Reif A, Hardison NE, Peters EJ, Havener TM, et al. (2011) Identification and replication of loci involved in camptothecin-induced cytotoxicity using CEPH pedigrees. *PLoS One* 6: e17561.
4. Peters EJ, Motsinger-Reif A, Havener TM, Everitt L, Hardison NE, et al. (2011) Pharmacogenomic characterization of US FDA-approved cytotoxic drugs. *Pharmacogenomics* 12: 1407-1415.
5. Huang RS, Duan S, Kistner EO, Hartford CM, Dolan ME (2008) Genetic variants associated with carboplatin-induced cytotoxicity in cell lines derived from Africans. *Mol Cancer Ther* 7: 3038-3046.

6. Duan S, Bleibel WK, Huang RS, Shukla SJ, Wu X, et al. (2007) Mapping genes that contribute to daunorubicin-induced cytotoxicity. *Cancer Res* 67: 5425-5433.
7. Brown CC, Havener TM, Medina MW, Auman JT, Mangravite LM, et al. (2012) A genome-wide association analysis of temozolomide response using lymphoblastoid cell lines shows a clinically relevant association with MGMT. *Pharmacogenet Genomics* 22: 796-802.
8. Watson VG, Hardison NE, Harris T, Motsinger-Reif A, McLeod HL (2011) Genomic profiling in CEPH cell lines distinguishes between the camptothecins and indenoisoquinolines. *Mol Cancer Ther* 10: 1839-1845.
9. Watters JW, Kraja A, Meucci MA, Province MA, McLeod HL (2004) Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proc Natl Acad Sci U S A* 101: 11809-11814.
10. Paolini GV, Lyons RA, Laffin P (2010) How desirable are your IC₅₀s? A way to enhance screening-based decision making. *J Biomol Screen* 15: 1183-1193.
11. Coval ML (1970) Analysis of Hill interaction coefficients and the invalidity of the Kwon and Brown equation. *J Biol Chem* 245: 6335-6336.
12. Brown C, Havener TM, Everitt L, McLeod H, Motsinger-Reif AA (2011) A comparison of association methods for cytotoxicity mapping in pharmacogenomics. *Front Genet* 2: 86.
13. Colquhoun D (2006) The quantitative analysis of drug-receptor interactions: a short history. *Trends Pharmacol Sci* 27: 149-157.
14. Beam AL, Motsinger-Reif AA (2011) Optimization of nonlinear dose- and concentration-response models utilizing evolutionary computation. *Dose Response* 9: 387-409.
15. R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
16. Brown CC, Havener TM, Medina MW, Krauss RM, McLeod HL, et al. (2012) Multivariate methods and software for association mapping in dose-response genome-wide association studies. *BioData Min* 5: 21.
17. Brown CC, Havener TM, Medina MW, Krauss RM, McLeod HL, et al. (In Press) R: Genome-wide Association and Pharmacological Profiling of 29 Anticancer Agents Using Lymphoblastoid Cell Lines. *Pharmacogenomics*.