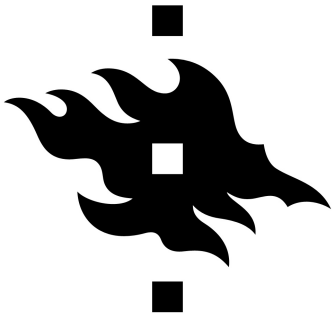

UNIVERSITY OF HELSINKI

Estimation of bias in dose-response curve fitting and experimental strategies to its reduction

Master's thesis

Master's Programme in Life Science Informatics

Diogo Dias

June 7, 2022

Supervisor(s): Jing Tang (Principal Investigator, Department of Mathematics and Statistics, University of Helsinki), Alina Malyutina (Doctoral Researcher, University of Helsinki)

University of Helsinki, Finland

Faculty of Medicine

... To the memory of my grandmother Clara and grandfather Artur

Acknowledgments

I want to first thank Sampsa Hautaniemi for being my Master's teacher tutor and to introduce and motivate me to pursue an academic career in cancer research in the Research Program in Systems Oncology (ONCOSYS) at the University of Helsinki.

Secondly, to Jing Tang. Jing is an amazing person and principal investigator. He gave me the opportunity and believed in me very early during my MSc studies to conduct scientific research in his research group *Network Pharmacology for Precision Medicine*. Without it, my path would probably not be as successful and I am extremely thankful for all his guidance.

In third, to Alina Malyutina. Alina is a very special person that I had the privilege to meet and work alongside with. It was an unique experience upon which I grew both as a person and as a future academic researcher. I will carry on with me all of her teachings and I wish nothing less than success and happiness for her. I also want to thank both Jie Bao and Johanna Eriksson for performing the drug sensitivity experiments for the project from Jing Tang's research group.

Lastly, to my mother, father and brother for the unconditional love and incredible support throughout this journey. Travelling to a new country and environment to pursue my dreams, no matter how important and big they are, it is still a big leap and their support was indispensable.

This research was supported by the European Research Council (ERC) starting grant, No.716063 (DrugComb: Informatics approaches for the rational selection of personalized cancer drug combinations), the Academy of Finland, No.317680, the Sigrid Juselius Foundation, the cancer Foundation Finland, the Instrumentarium Science Foundation, and the Päivikki and Sakari Sohlberg Foundation.

Abstract

One of the biggest hurdles in cancer patient care is the lack of response to treatment. With the support of high-throughput drug screening, it is nowadays feasible to conduct vast amounts of drug sensitivity assays, aiding in the identification of sensitive and resistant samples to chemical perturbations. In an oncology setting, drug screening is the process by which patient cells are examined experimentally for response and activity to distinct drugs and analysed via dose-response curve fitting. However, the ability to reproduce and replicate with high confidence drug screening outcomes proved to be a challenge that needs to be addressed. Inefficient experimental designs, lack of standard protocols to control both biological and technical factors in such cell-based assays are at the core of a steep influx of experimental biases. Hence, additional endeavour has to be carried out to provide less biased estimations of drug effects. This thesis work focuses on reducing erroneous inferences (i.e., bias) from dose-response data in the curve fitting step, thereby improving the reproducibility of drug sensitivity screening through efficient dose selection. A novel two-step experimental design is introduced which significantly improves the estimation of dose-response curves while keeping the amount of cellular and chemical materials feasible.

Thesis abbreviations

AA Activity area

ABC Area between the two drug dose-response curves

AIC Akaike information criterion

AML Acute myeloid leukemia

ATP Adenosine triphosphate

AUC Area under the curve

BAO BioAssay ontology

$BFGS$ Broyden-Fletcher-Goldfarb-Shanno

BIC Bayesian information criterion

$CCLE$ Cancer cell line encyclopedia

$COSMIC$ Catalogue of somatic mutations in cancer

CSS Drug combination sensitivity score

$CTRP$ Cancer therapeutics response portal

CVD Coefficient of variability in difference

DIP Drug-induced proliferation rate

$DMSO$ Dimethyl sulfoxide

DR Drug-response

$DSRT$ Drug sensitivity and resistance testing

DSS Quantitative drug sensitivity scoring

$ELISA$ Enzyme-linked immunosorbent assay

FGF Firoblast growth factor

$GDSC$ Genomics of drug sensitivity in cancer

GR Normalized growth rate inhibition

GSC Glioblastoma stem cell

$HDAC$ Histone deacetylase

HSA Highest single agent

HTS High-throughput screening

$L-BFGS-B$ Limited-memory Broyden-Fletcher-Goldfarb-Shanno

LRT Likelihood ratio test

$MCLP$ MD Anderson cell lines project

MM Multiple myeloma

NDR Normalized drug responses

NMF Non-negative matrix factorization

$recGBM$ Recurrent glioblastoma

RI Relative inhibition

$sDSS$ Selective drug-sensitivity scores-response

$SSMD$ Strictly standardized mean difference

TAS Target addiction scoring

$VEGF$ Vascular endothelial growth factor

ZIP Zero interaction potency

Contents

Chapter 1

Introduction

Assessment of drug-response (DR) relationship is at the core of understanding the mechanisms of action, offering breakthroughs in cancer research that ultimately lead to effective therapies (Malani et al. 2021; Pemovska et al. 2013). With the support of high-throughput screening (HTS), it is feasible to screen thousands of chemical compounds that target multiple signalling pathways or metabolic states for a variety of cancer cell lines (Manuvakhova et al. 2011; Pietarinen et al. 2015). This swift development has fostered a deeper understanding of cellular responses to chemical perturbations. In a drug sensitivity experiment, cell-based assays are commonly performed in two-dimensional (2D) cell cultures by which cancer cell lines are examined for response in activity (Jakštys et al. 2015; Markossian et al. 2021). In this thesis work, drug sensitivity data is collected and analysed in multiple myeloma (MM) cancer cell lines. Cell viabilities with their respective drug concentrations are fitted via a four-parameter logistic function. Then, using different drug sensitivity metrics, drug effect can be quantified. Therefore, it is crucial to obtain the estimates as accurate as possible to increase experimental reliability. Notwithstanding, one of the grand challenges is the lack of reproducibility and replicability in the experimental phase. For

example, it has been reported that drug sensitivity experiments on the same cancer cell lines, using equal drug concentration ranges from different study centers produced distinct DR curves within the same and across different centers (Niepel et al. 2019). Optimizing such assays with proper experimental design is crucial to obtain reproducible results (Larsson et al. 2020). Despite the problem being highlighted in several studies (Consortium et al. 2015; Haibe-Kains et al. 2013; Haverty et al. 2016), guidelines to generate DR data in a more reproducible manner have not been widely explored nor proposed for that matter. Taken together, a novel experimental design is needed to improve both the reproducibility in DR assays and the accuracy of the estimated drug effect.

Chapter 2

Literature review

2.1 Drug sensitivity testing

Historically, Yoshida et al. (1975) investigated drug potency and efficacy in cultured cells by means of 5-fluorocyclocytidine detection, paving an avenue for further advancements in pharmacological testing. Drug sensitivity and resistance testing is an essential step of pharmacological studies to investigate drug responses at individual cancer samples. *Ex vivo* drug sensitivity, once reliably measured and quantified, is able to suggest safe and effective personalized therapies in the form of molecular tumor boards (Malani et al. 2021).

2.1.1 High-throughput screening

After the onset of the NCI60 cell-line study (Shoemaker 2006) which aimed at facilitating drug discovery by offering HTS screening data, a report by Weinstein (2012) further reinforced importance in the search for novel treatments. With the aid of HTS assays, a wide array of drug testing is conducted in multi-well plates (i.e., *96*-, *384*- or

1536 -wells) depending on the available laboratory material and research question that are susceptible to both technical and biological sources of variation. After target identification and decide whether to activate or inhibit a specific molecule or pathway, the assay is set up and HTS is optimized for as little bias in the experiment as possible. A typical drug sensitivity experiment is represented below upon which first, blood samples are taken from a patient and cancer cells are separated for testing and diluted. Then, the assay is performed in cell culture plates where a selected drug of interest is present on each well with cancer cell lines at a different concentration. Drug sensitivity profiling is then computed in downstream analysis and a DR relationship is inferred. This type of workflow is widely adopted in current pharmacological studies taking advantage of large oncology libraries (Pemovska et al. 2013; Yadav et al. 2014). Below is illustrated a typical drug screening experiment (see Figure 2.1).

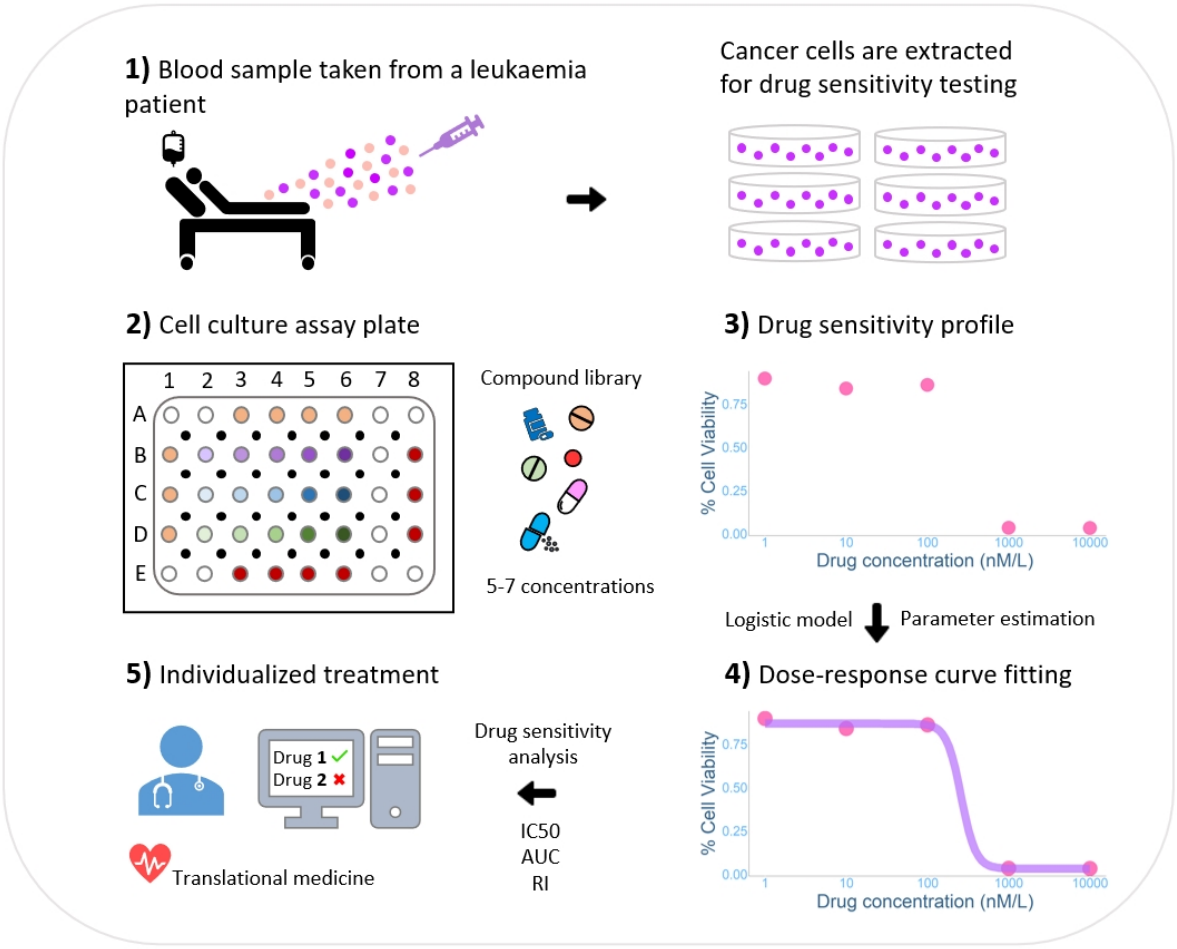

Figure 2.1: Workflow of a typical *ex vivo* drug sensitivity assay pipeline for leukaemia.

2.1.2 Public resources for the discovery of anticancer compounds

During mid 2000's, Posner (2005) highlighted potential obstacles and impact of HTS assays in the pharmaceutical industry. To address such limitations, BioAssay Ontology (BAO) was developed (Visser et al. 2011) as an attempt to standardize HTS experiments and was further tested for assay notation and comparative analysis of different methods (Zander Balderud et al. 2015). This BAO has facilitated the development of chemical

biology databases such as the widely used drug discovery resource ChEMBL (Mendez et al. 2019), the key resource for biomedical research PubChem (Kim et al. 2021), and DrugTargetCommons (DTC) (Tang et al. 2018) for drug repurposing and detection. To offer a tailored and robust open-access data portal for drug combination screening studies for a large number of cancer cell lines, DrugComb (Zagidullin et al. 2019) was developed and more recently updated with more comprehensive screening data as well as machine learning algorithms to tackle several diseases (Zheng et al. 2021). Additionally, to aid in the rational design of synergistic drug combinations via pairwise evaluation in cell-based assays, SynergyFinger (Ianevski et al. 2017) was developed and further refined by including additional curve fitting models, visualization alternatives and statistical tools (SynergyFinger Plus) (Zheng et al. 2022).

Around a decade ago, two leading pharmacogenomic HTS studies on cancer cell lines CCLE (Barretina et al. 2012) and GDSC (Yang et al. 2012) accelerated pharmacological discovery of candidate drug targets and molecular therapeutics, offering high quality genomic data but less concordant drug sensitivity data. Additional large and important databases comprising genomic, transcriptomic and drug sensitivity data were developed afterwards. For instance, COSMIC (Forbes et al. 2017; Tate et al. 2019), being the largest and the most detailed database for somatic mutations in cancer, has become the state-of-the-art resource for mutational signature analysis (Alexandrov et al. 2020) assisted by advanced machine learning algorithms such as the non-negative matrix factorization (NMF) (Lee et al. 2000). Moreover, CTRP (Basu et al. 2013) was developed to facilitate the development of novel therapeutic strategies with their pharmacogenomic data taking into account genomic confounding factors of cell lines, and further expanded to the second version CTRP v2 (Seashore-Ludlow et al. 2015). Li et al. (2017) published the MCLP data portal by adding protein expression data of more than 200 cancer-related proteins for 600 cancer cell lines. For drug combinations, a

non-biased dataset involving screening of 22,737 experiments, 583 drug combinations in 39 distinct cancer cell lines was presented (O'Neil et al. 2016). NCI-ALMANAC (Holbeck et al. 2017) dataset was developed to screen and assess more than 5000 drug pairs against 60 tumor cell lines and is a common resource in machine learning applications (Julkunen et al. 2020). Similarly, a couple of years back, to enable a robust comparison on variational datasets, *Smirnov* and colleagues published PharmacoDB (Smirnov et al. 2018), focusing on improving reproducibility of cell viability assays (Larsson et al. 2020). For a more in-depth survey, a recent review analyzed and summarized 102 public databases based on both quality and category (Tanoli et al. 2021). Altogether, drug sensitivity profiles have shown great potential, especially if such outcomes are reproducible. Table 2.1 displays major public resources.

Resource	**Acessibility**	**Reference**
CCLE	*CCLE collection*	(Barretina et al. 2012)
GDSC	*GDSC website*	(Yang et al. 2012)
CTRP/v2.0	*CTRP portal*	(Seashore-Ludlow et al. 2015)
O'Neil	*O'neil article*	(O'Neil et al. 2016)
MCLP	*MCLP web app*	(Li et al. 2017)
NCI-ALMANAC	*NCI-ALMANAC navigation*	(Holbeck et al. 2017)
PharmacoDB	*PharmacoDB documentation*	(Smirnov et al. 2018)
ChEMBL	*ChEMBL database*	(Mendez et al. 2019)
PubChem	*PubChem homepage*	(Kim et al. 2021)

Table 2.1: Major pharmacological resources for drug sensitivity studies.

2.1.3 Measuring cellular response

The classical way of counting viable cells is visual inspection under a microscope. Such a process can be tedious and inefficient (Pierre 2002). In the past, tetrazolium salts were frequently used to measure activity of cells and enzymes (Slater et al. 1963). The first cell viability assay with a color reagent was performed using tetrazolium salt measuring living cells (Mosmann 1983). Additional efforts were carried out to determine numbers of cultured cells such as the CyQUANT assay (Jones et al. 2001) yielding a fluorescent-like signal upon DNA binding, and the reagent *Alamar Blue* dye through color shift determined via enzyme-linked immunosorbent assay (ELISA) (Ahmed et al. 1994). This measurement technique was applied to cytotoxic studies in mammalian cells (O'brien et al. 2000) and further improved for hepatocytes and liver cell lines (McMillian et al. 2002). Further efforts were made to detect cell death through detection of differential proteolytic activities (Niles et al. 2007). However, the latter techniques assume that every cell within a population reacts evenly to a drug which is usually not the case (i.e., cell heterogeneity). Finally, adenosine triphosphate (ATP) concentrations are among the most commonly used cell viability analyses using the CellTiter-Glo reagent (Tolliday 2010).

Nonetheless, all of the previously mentioned cellular measurement techniques may produce discordant results applied to the same cell-line, compound, and laboratory material. This phenomena was observed around a decade ago where cell numbers were discordant using ATP and tetrazolium-reduction assays (Chan et al. 2013; Quent et al. 2010). In principle, in HTS assays, the detection reagent contains an enzyme that reacts with ATP (*luciferase enzyme*) yielding light for direct measurement of cell viability in cancer cell lines (Zheng et al. 2013). The weaker the signal, the fewer viable cells are present in the growth medium, suggesting that the drug has inhibited cancer cells more

efficiently. Such luminescence or light signal is quantified with a microplate reader at appropriate excitation wavelengths. This measurement is typically performed 72 or 92 hours after plate incubation (Jaaks et al. 2022). The percentage cell viability values are normalized in relation to both positive and negative controls and subjected to drug sensitivity data analysis pipelines.

2.2 Reproducibility challenge for drug sensitivity assays

Inconsistent replicability of drug sensitivity assays pose an obstacle in accurate estimation of drug effects. Recently, multiple studies have identified specific factors of irreproducibility in drug sensitivity experiments. In 2019, Niepel et al (Niepel et al. 2019) reported the results of five research centers performing drug sensitivity experiments on the same cell line, culture media, and chemical compound (Figure 2.2, **A**). To increase the reliability of the results, each center used various biological and technical replicates. However, there have been observed big variations in the resulting DR curves within and across the centers, suggesting procedural variation and low reproducibility (Figure 2.2, **B**).

Importantly, a recent study highlighted the importance of minimizing the effects of experimental confounders in cell viability assays such as cell culture conditions, drug concentration and treatment time, DR metrics, and medium protocols for assay optimization (Larsson et al. 2020). However, practical strategies to generate reproducible DR curves to a greater extent have not been developed.

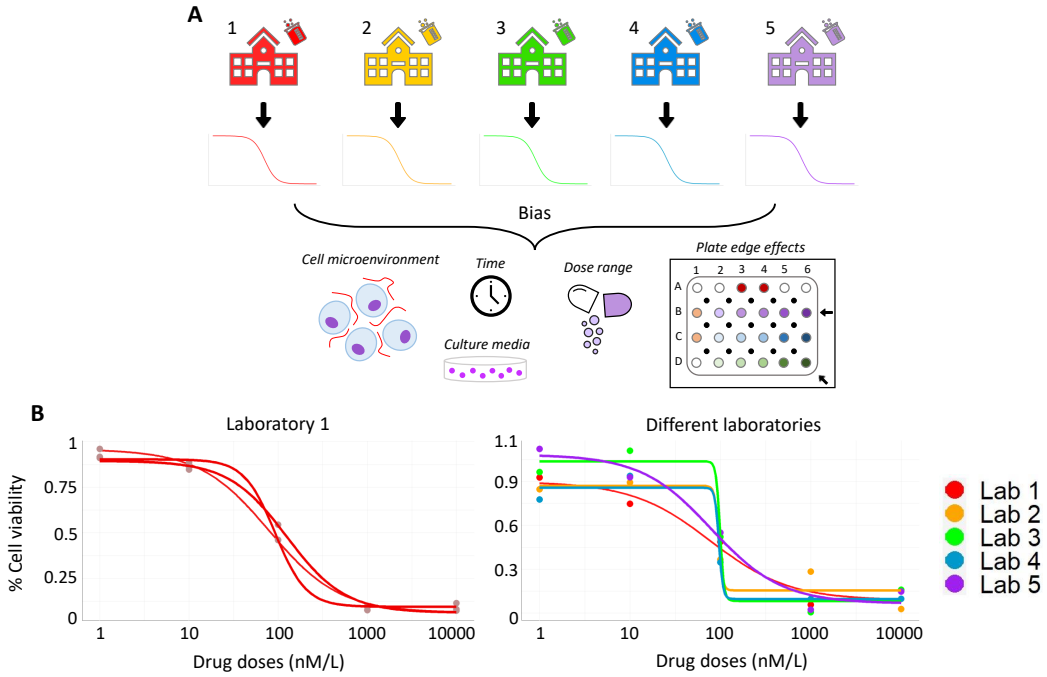

Figure 2.2: Comparison of the resulting dose-response curves within one and across different labs. **A.** Five research labs perform drug sensitivity experiments using the same experimental settings to avoid technical and biological biases. **B.** Comparison of the resulting dose-response curves within and across distinct labs.

2.2.1 Lack of reproducibility in large-scale pharmacological datasets

The CCLE and GDSC databases consist of drug sensitivity and genomic data. Nonetheless, the lack of standardized protocols and inconsistent drug effects pose a huge hurdle in drug discovery. The following year after the introduction of CCLE and GDSC datasets, Haibe-Kains et al. (2013) carried out a comparative analysis of common drugs that accounted for variation and found that outcomes from both resources were not similar (i.e., a potent and effective drug in a specific cell line within one dataset did not show the same effect on the other). This study raised important questions on whether drug sensitivity data from these two datasets were reproducible. It is widely assumed

that because of this study, the CCLE and CDSC authors launched an analytical analysis to demonstrate that reproducibility was true (Consortium et al. 2015). Haibe-Kains and colleagues were not satisfied by the analysis performed on both CCLE and GDSC datasets in 2015 which lead to an extra study and found once again that measures of drug sensitivity were not consistent (Safikhani et al. 2016a,b). It appears that the inconsistency of the results might depend on the chosen DR metric, highlighting the importance of standardized approaches to assess the reproducibility of drug sensitivity. Figure 2.3 lists major studies focusing on reproducibility. Ideally, using robust DR metrics, consistency can be improved. However, no standardized approaches have been proposed.

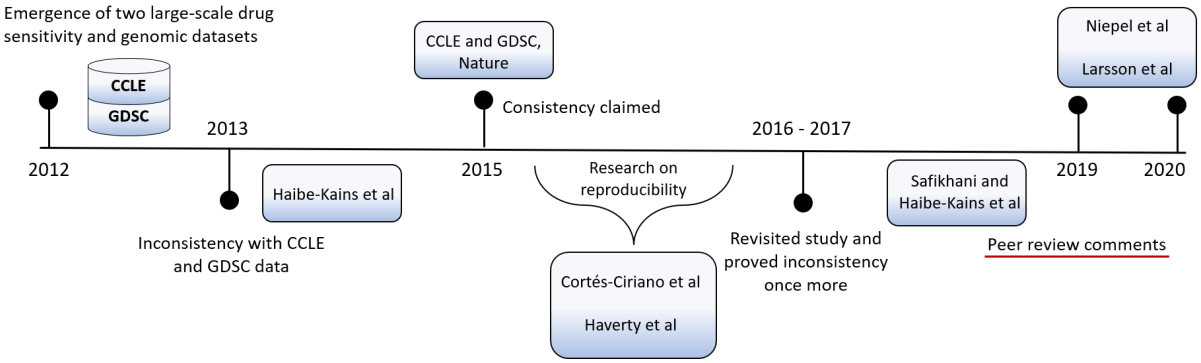

Figure 2.3: Reproducibility studies in CCLE and GDSC datasets.

2.2.2 The impact of biological and technical drivers of variability in drug sensitivity studies

A comprehensive characterization and understanding of molecular features is needed to reduce batch effects and bias originated by both technical and biological phenomena. It is alarming that in the industrial community, failure to confirm important findings in

scientific journals have been reported (Baker 2016). In addition, a two-year study (2008 - 2010) attempted to understand the lack of success in Phase II clinical trials regarding several chemical compounds (Graul et al. 2009, 2010). Then, Arrowsmith (2011) pointed that a collaborative effort in the industry is needed for reproducible findings in clinical trials. A crucial report on the open investigation of the reproducibility in oncology stated that *"replication is central to the progress of science: if others cannot reproduce the evidence backing a scientific claim, then the claim loses status as scientific knowledge"* (Errington et al. 2014). One believes that vigorous and standardized rules in the field of science are required for efficient reproducibility and replicability of scientific claims from several fields. After the resurgence of large-scale pharmacological and pharmacogenomic datasets and the thorough global discussion therein on the issue of inconsistent outcomes, an essential study on providing proper guidelines for the reuse of such resources took place (Wilkinson et al. 2016) (i.e., the *FAIR* Guiding Principles).

Indeed, for higher efficiency of HTS assays, two main sources of variability need to be taken into account when designing a drug sensitivity experiment for acceptable results (Niepel et al. 2017). The biology of DR behaviours can be extremely difficult to control, especially in independent experiments considering that standardization of cell type, growth medium composition, cell density, plate edge effects, drug dosage and type, freezers, and incubation time will certainly differ from laboratory to laboratory. Furthermore, evolutionary dynamics of drug resistance (Zur Wiesch et al. 2011) have been reported to affect DR on cultured cells (Ramirez et al. 2016). Accordingly, bias in cell-based assays was shown to be derived from the day and time at which the measurements were taken (Harris et al. 2016) which indicates another important source to be considered (i.e., *time-dependent bias*).

2.2.3 Quality control in HTS assays

In HTS assays, one single error in the automation process can jeopardize the whole experiment. Celltiter-Glo, the most commonly used quantification method in a drug sensitivity experiment, is also susceptible to errors which result in inconsistent DR relationships. Important drivers of variability that are usually considered are plate edge effects due to uneven evaporation of the plate edges, robotic liquid handling, laboratory material and environment of the experiment (Bushway et al. 2010; Root et al. 2003). Assessing HTS technical errors via quality control is usually performed with the use of Z-factor (Mpindi et al. 2015) being determined with dimethyl sulfoxide (DMSO) treated positive and negative controls (Jaaks et al. 2022). These screening controls are crucial in evaluating plate edge effects and possible technical bias in modern experiments, however, they are often not enough to increase the quality of the experiment by themselves. Further, the author responsible for Z-factor (Zhang et al. 1999), in 2007, published new quality control (QC) methods such as the strictly standardized mean difference (SSMD) and the coefficient of variability in difference (CVD) to aid in HTS assays (Zhang 2007).

Bioinformatic tools are key to optimize HTS assays through quality control, parameter estimation, and informative visualisations. Hence, experimental data need to first be collected, systematized, measured, and visualized in a proper way by means of advanced tools. Recently, integrated quality control and data analysis applications for HTS have been developed and provided as open-source tool (Ianevski et al. 2017; Potdar et al. 2020). These applications are crucial to further identify individual drug resistance and sensitivity patterns for personalized patient care as well as for rational designs of drug combinations that are able to tackle resistant mechanisms of cancer cells and enhance treatment efficiency. Lastly, Mpindi et al. (2015) studied key factors in HTS experiments and emphasized the importance of both scattered plate layout of controls

and the use of Loess method for normalization, thereby reducing place edge effects and generate accurate DR curves.

2.3 Dose-response curve fitting

In cancer research, DR relationship inference is at the core of understanding drug responses. In a typical drug sensitivity experiment, each cell viability data point with respect to its dose concentration value is fitted to a four-parameter logistic model (Findlay et al. 2007; VØlund 1978) which usually displays a sigmoidal curve (Vis et al. 2016). The independent variable of such a model is the drug concentration and the dependent variable is the cellular viability. Hence, each curve has its parameters which can be used to determine multiple drug sensitivity metrics (Yadav et al. 2014). The fitting of a model is achieved by estimating the parameter values that best represents its true behaviour. However, when the logistic model does not fit the data properly (i.e., it does not represent a sigmoidal relationship), the IC50 (crucial region when cells are inhibited by half) is likely ill-defined. In such cases, non-parametric models might be needed to calibrate cellular responses using *Spline*, *Monotonic* or *Bayesian* approaches (Amiryousefi et al. 2021). The DR curve fitting experienced a high peak of development over the last years due to the evolution of the open-source environment R where new packages and algorithms within surfaced that allow for quantitative analyses of cellular dynamics. Contemporary state-of-the-art packages based on specific algorithms and tools include **drc** (Ritz et al. 2015), **nplr** (Commo et al. 2016), **drda** (Malyutina et al. 2021), and **DoseFinding** (Bornkamp et al. 2010). Alternative resources include *nlstools* for hormetic curve fitting (Baty et al. 2015), *MCPMod* for the design and analysis of dose finding trials (Bornkamp et al. 2009), and *grofit* for the fitting of growth curves under different biological scenarios (Kahm et al. 2010). Throughout this thesis, the **drc**

package will be utilized.

The common model to fit a DR relationship curve is the 4-parameter generalized logistic function (Equation 2.1).

$$y = \beta + \frac{\alpha - \beta}{1 + e^{-\eta(x-\phi)}} \tag{2.1}$$

Where x are the drug doses in raw scale, β and α stand for lower and upper bounds of the response respectively, η determines the potency of a drug (i.e., how fast or slow cells are being inhibited - *slope*), and ϕ (or *IC50*) is defined as the concentration at which the drug inhibits 50% of cells. Providing that the ϕ is reached with the initial tested doses, suggests drug-sensitive cells. On the contrary, in cases which the majority of test doses do not inhibit half of the cell population, it suggests a weak drug response. In the latter case, it is common to use synergistic combinations to achieve higher treatment efficacy (Szwajda et al. 2015; Tang et al. 2013). Alternative models for DR curve fitting are, for instance, the 2-parameter logistic model, the 5-parameter function (*or Richard's curve*) (Richards 1959), and *Emax* (Macdougall 2006).

2.3.1 Algorithms to infer dose-response relationship

Currently, the **drc** package is one of the most common resources for DR curve fitting in drug sensitivity assays. This package contains several functions which have flexible and versatile model fitting and evaluation functions, to properly capture the *true* cellular activity, allowing users to choose a nonlinear model for curve fitting from an ample range of sigmoid functions. In addition, **drc** applies quasi-Newton methods such as the Broyden-Fletcher-Goldfarb-Shanno (BFGS) (Fletcher 2013) and the limited-memory BFGS (L-BFGS-B) that has shown improved numerical performance compared to

previous optimization algorithms (Liu et al. 1989). In addition, certain functions within the package are able to be extended by other R packages such as *medrc* (Gerhard et al. 2017) and *nlme* (Pinheiro et al. 2007).

drda is a fairly recent package developed and proved high performance in the residual sum squares paradigm which is able to provide accurate and consistent results to a greater degree (Malyutina et al. 2021). Instead of applying a numerical optimization routine, **drda** implements a new Newton method with a trust region through conjugate gradient for large-scale optimization problems (Steihaug 1983). This overcomes the main limitations of previous packages using quasi-Newton methods to minimize non-convergence problems using logistic models for inference. While other algorithms employ a single initialization step, **drda** doubles the steps to optimize algorithm convergence. In addition, the algorithm relies on the analytical gradient and Hessian formulas instead of numerical approximations in problematic scenarios to assure proper convergence. Furthermore, it computes confidence intervals for the whole DR curve using multiple comparisons tests correction such as likelihood ratio test (LRT) through Akaike information criterion (AIC) (Akaike 1974) and Bayesian information criterion (BIC) (Schwarz 1978).

The **DoseFinding** package was designed for phase II clinical trials providing functions for the design and analysis of DR experiments. Similarly to **drc**, **DoseFinding** also takes advantage of the quasi-Newton algorithm BFGS and enforces boundary constraint to non-linear models parameters. Bornkamp et al. (2011) studied the challenges of dose-finding studies (i.e., model uncertainty and variability in the parameter estimates) and proposed adaptive designs in clinical studies under different scenarios. Later on, using general parametric models, they also developed a robust framework for model comparison in dose-finding studies (Pinheiro et al. 2014).

The **nplr** package provides generalized logistic functions based on the 5-parameter logistic regression model utilizing classic Newton approaches (i.e., Newton-Raphson method) employed in the *nlm()* function of the package. Furthermore, **nplr** provides users metrics to assess the area under the curve and confidence intervals for estimated drug doses. Albeit fast, it has the disadvantage of providing poor DR inferences due to improper convergence approximated numerically.

2.3.2 Assessing parameter effect on the shape of a curve

It is essential to evaluate the behaviour of the 4-parametric logistic model (Equation 2.1) when varying the parameters (Figure 2.4).

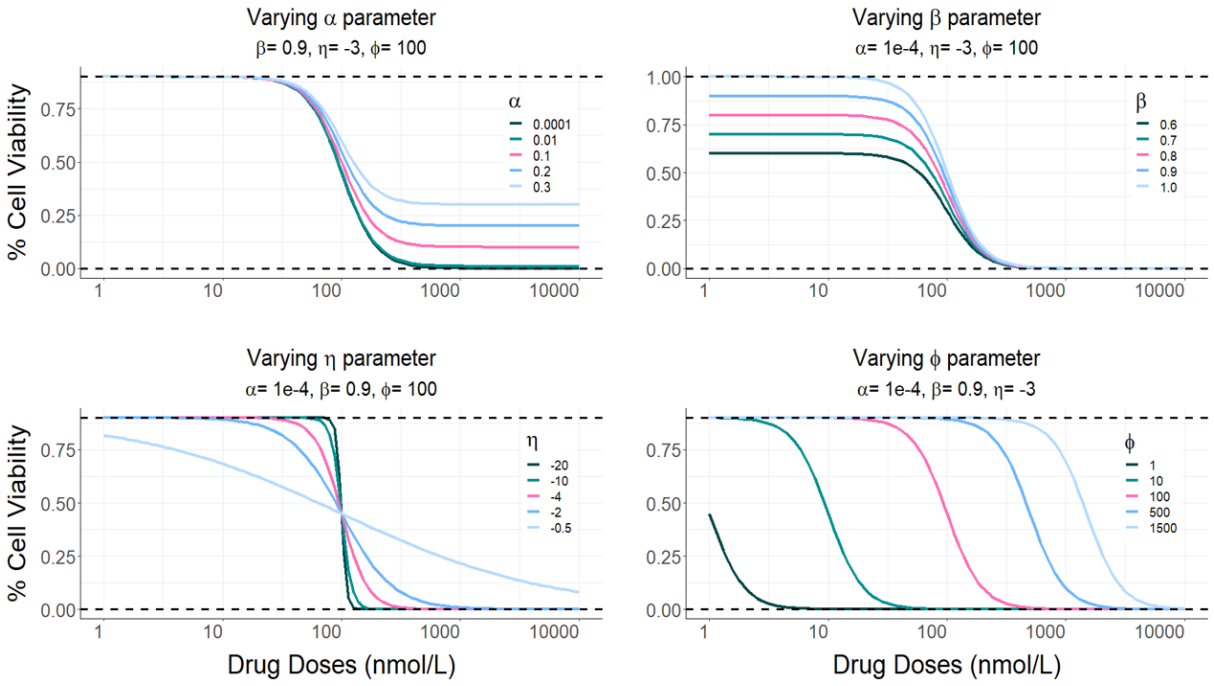

Figure 2.4: Parameter assessment of the four-parameter logistic function in inhibitory scenarios where slope is negative. On each plot, a specific parameter of the model is varying while others are fixed as highlighted for each scenario.

Varying only the β or α (curve bounds) parameters of the model only alter the minimum and maximum response of cells in the presence of chemical compounds, respectively. In cell viability terms, the maximum response indicates the total cell population and the minimum indicates the point at which cells are completely inhibited. Fluctuating the η parameter demonstrates how cells react to a particular chemical perturbation and how fast cells are being inhibited from one concentration level to the next. The ϕ parameter is commonly used to quantify the potency of a drug. The lower the ϕ, the more efficiently a compound inhibits cell viability. On the contrary, the bigger the ϕ, the more doses are needed to inhibit cells which displays less efficacy. *A priori*, it appears that η and ϕ parameters of the model are of interest and worthy of exploration.

2.3.3 Curve fitting summary

In order to determine the optimal parameters, all of the above packages except **drda** employ iterative Newton approaches extensively (Nocedal et al. 2006). Inconsistent results of curve fitting when applied to similar pharmacological data from HTS experiments or from global resources with the exact same model proved to be a challenge in the discovery of effective compounds. Therefore, not only this thesis will evaluate **drc** over a wide spectrum of parameters of the model (Equation 2.1) using massive simulations, but will also hypothesise, propose, and validate efficient solutions for higher replicability and reproducibility of drug sensitivity assays through different dose selection processes.

2.4 Identification of drug candidates

Analysis of sensitive and resistant cells in response to chemical perturbations have usually used the IC50 value as the main metric (Dong et al. 2015; Fallahi-Sichani et al.

2013). Attempts to optimize the accuracy of the IC50 estimation through multilevel modelling were reported (Vis et al. 2016). An alternative to IC50 is the half-maximal effective concentration (EC50), being the required dose to experience 50% of the drug effect. Fundamentally, if inhibition is measured, the IC50 is used at which point inhibits 50% of cell population compared to untreated controls. In contrast, if viability is measured, the EC50 metric is adopted. In HTS assays, cancer cell lines are usually exposed under five different concentrations in a log10 scale (e.g., *1*, *10*, *100*, *1000*, *10000* nmol/L) (Pemovska et al. 2013). Drug scoring in individual patients is a crucial step in drug discovery. Hence, various metrics have been developed over the years to detect drug efficacy and potential drug-target candidates.

2.4.1 Quantification metrics of drug sensitivity

DR quantification is an essential process at the forefront of proposing effective treatment in personalized medicine. Scoring *ex vivo* drug sensitivity data allows for a broader view on a systems-wide level by understanding how the cells behave and, potentially, what genetic modifications the cancer cell rely on. Hafner et al. (2016) developed a new metric called GR (Growth rate) to measure cell sensitivity to cancer drugs and showed that classical pharmacological metrics fail to obtain consistent outcomes mostly due to differences in cellular proliferation. The CCLE study (Barretina et al. 2012) developed the activity area (AA) to identify the efficacy and potency of 24 compounds in a large collection of cell lines. Additionally, the quantitative drug sensitivity scoring (DSS) and selective drug-sensitivity scores-response (sDSS) based on the area under the curve (AUC) (Ekstrøm 2020), showed performance in (Yadav et al. 2014) for patients with acute myeloid leukemia (AML). In events such that positive and negative controls are available, relative inhibition (RI) can be considered as a proper metric and computed for each dose plate (Malyutina et al. 2019). Curiously, Fallahi-Sichani et al. (2013) used a

combination of the (*1*) slope parameter of DR curves, (*2*) AUC, and (*3*) the upper bound (or maximum response) to assess and prove the strength of a multiparametric approach in drug sensitivity quantification. Using both the CCLE and GDSC data, consistency of cell viability data was accomplished by computing the area between two DR curves (ABC) (Safikhani et al. 2016b). Remarkably, an interesting *in vitro* study proposed the drug-induced proliferation (DIP) rate (Harris et al. 2016) as a quantification metric to remove temporal bias in antiproliferative compound activity on cells. Their study highlighted the importance of both cellular growth and lag in drug effect on cells as potential causes of time-dependent bias. Recently, to reduce erroneous inferences in interpreting drug sensitivity data, normalized drug response (NDR) (Gupta et al. 2020) was developed and benchmarked via simulations. This new metric also makes use of positive and negative controls in HTS experiments and it was shown to improve consistency and accuracy of dose-responses.

Multi-targeted treatments such as drug combinations have shown potential in treating complex diseases (Adam et al. 2018; Sun et al. 2017). Malyutina et al. (2019) developed a drug combination sensitivity score (CSS) to score drug pair potency to facilitate the discovery of synergistic drug effects. Bliss independence is a traditional model where two drugs A and B act in an independent fashion, thus expecting their effect to be amplified (Bliss 1939). Other synergy scoring metrics include the highest single agent (HSA) (Berenbaum 1989), Loewe additivity (Loewe 1953), and the zero interaction potency (ZIP) (Yadav et al. 2015). Currently, these drug combination metrics are provided in the *SynergyFinder Plus* standalone web application (Zheng et al. 2022).

2.5 Applications of drug sensitivity assays in biomedicine and biomarker discovery

For optimal anticancer therapy, HTS allows for large-scale drug testing able to aid in DR biomarker discovery. Such automation is extensively used in cancer research. In just a couple of days, a large collection of chemical compounds is screened against patient-derived primary samples or cell lines (Collignon et al. 2020; Lin et al. 2020). Examples of this type of assay approach have been used to understand DR variability and patient stratification in AML patients (Majumder et al. 2017; Pietarinen et al. 2015). Drug sensitivity and resistance testing (DSRT) (Malani et al. 2017; Malani et al. 2021; Pietarinen et al. 2017) demonstrated its power in DR biomarker discovery (Pemovska et al. 2015). Moreover, it has been shown that *in vitro* drug sensitivity was able to predict oncogenic signaling pathways (Tyner et al. 2013). A computational method termed target addiction scoring (TAS) (Jaiswal et al. 2019) was able to convert DR profiles into drug target signatures allowing for compound ranking. Towards addressing intertumoral heterogeneity in recurrent glioblastoma (recGBM) cancer cells, Skaga et al. (2019) applied DSRT to glioblastoma stem cell (GSC) cultures via DSS metric and one-way ANOVA statistical testing to observe significantly differences in drug sensitivity across cell cultures.

Drug sensitivity profiling goes beyond haematological diseases. In breast cancer, drug screening is able to identify candidate predictors of response (Heiser et al. 2012). Correlation between drug sensitivity and molecular features revealed around 250 genes associated with response to treatment (Kuo et al. 2009). Representative drug sensitivity applications in biomedicine are illustrated in Table 2.2.

Data	Description	Reference
Ex vivo drug sensitivity and molecular profiling	Drug sensitivity and resistance testing (DSRT) combined with genomic profiling was able to provide patient-specific therapy for 187 drugs in 18 patients with acute myeloid leukaemia.	*Pemovska et al., 2013*
Ex vivo drug sensitivity	Identification of novel drug candidates based on drug sensitivity scores and dose-response profiling for blast phase chronic myeloid leukemia using a total of 295 drugs. Effective drugs include VEGFR, NAMPT, and MEK inhibitors.	*Pietarinen et al., 2018*
Ex vivo drug sensitivity	This study focused on optimizing cell-based assays through the identification of both biological and technical factors using quality control metrics.	*Larsson et al., 2020*
Ex vivo drug sensitivity and targeted next-generation sequencing (tNGS)	Development of a promising tailored treatment strategy helped to provide personalized treatment for 28 acute myeloid leukemia (AML) patients in less than 21 days.	*Collignon et al., 2020*
Multi-omics ensemble (*ex vivo* drug sensitivity, exome and RNA sequencing)	Highlighted failure of genomic data alone to provide enhanced therapeutics in patients with acute myeloid leukaemia. A robust integration of omics data enabled the identification of a novel drug-response biomarker such as the overexpressed *IL15* gene in resistant samples.	*Malani et al., 2022*

Table 2.2: Examples of drug sensitivity applications in biomedicine.

Chapter 3

Aims of the study

The aim of this thesis is to tackle the reproducibility in drug sensitivity studies by reducing the bias that occurs during the drug effect estimation using dose-response curve fitting. A two-step experimental design is proposed as an efficient tool to plan the drug sensitivity assays.

The three main aims of this thesis are as follows:

1. Analyse if the proposed two-step design helps to reduce the dose-response bias and relative errors of common drug sensitivity metrics via computational simulations.

2. Confirm simulation findings through experimental validation on *OPM-2* cancer cells.

3. Final evaluation of the two-step experimental design (i.e., suitability for real experiments) and possible limitations of the approach.

Chapter 4

Materials and Methods

4.1 A two-step experimental design

A novel two-step experimental design that involves a pilot and validation experiment is introduced. At the core of the design, new drug doses are proposed by the algorithm based on the information from the pilot experiment. In the validation experiment, cell viabilities at new doses are measured and merged with the data from the pilot experiment. Finally, an accurate dose-response relationship can be inferred with the merged data.

This thesis will evaluate the curve fitting bias via large-scale simulations governed by different scenarios under the four-parametric logistic model (Equation 2.1). The two-step experimental design was initially built based on computational simulations and assumptions from cell-based assays. First, the strategy will be tested *in silico* and further validated using drug sensitivity assays with standardized protocols for OPM-2 cell lines.

Computational simulations and design evaluation: The new design will be

tested and evaluated via simulations in a large-scale framework with the use of the 4-parameter logistic function (Equation 2.1) over a large grid of parameters. DR curve fitting is computed by means of applying simulated drug sensitivity data to the **drc** package curve fitting pipeline. In the first step, involving a pilot experiment, drug doses and cell viabilities are first sampled similarly as in a real-life screening experiment from the true curve when all the parameters are fixed using normal distribution with mean (μ) centered at the true curve viability values and a standard deviation (σ) of 0.05. Then, new drug doses are proposed by the dose selection algorithm and selected for the second step, labelled as validation testing, based on information from the pilot experiment as input. Finally, both pilot and validation drug sensitivity data are merged, and a new curve is fitted with the merged dataset. Bias and relative errors of the main drug sensitivity metrics area under the curve (AUC) and ϕ (IC50) are then computed as the main evaluation metric as follows:

$$Bias = |\hat{x} - x| \tag{4.1}$$

Where $\hat{x}$ is the estimated dose-response curve values from the fitted drug sensitivity data, and x is the true dose-response curve values.

$$RE = |\frac{x - \hat{x}}{x}| \tag{4.2}$$

Where RE is relative error, $\hat{x}$ is the estimated either α, β, η, ϕ, *or* AUC, and x is the true computed values.

Drug screening experiments: OPM-2 cancer cell lines and drugs that cover distinct IC50 scenarios were used for experiments (see Table 4.2). In total, 8 cell culture

plates were used for the experimental validation (6 for the pilot experiment and 2 for the validation step). For the initial pilot assay, OPM-2 cells were subjected to doses at 5 concentrations in both log10 and uniform intervals with 1, 2, 3, and 5 replicates. For the validation step, was tested the addition of 3, 4, and 5 doses with 1 replicate. To assess the performance of the two-step experimental design denoted as the *ground truth*, OPM-2 cells were subjected to doses at 15 concentrations in log10 scale with a total of 3 replicates. Analysis of pilot assay plates ought to provide decisive insights for the selection of effective drug doses for the second round of experiments. Figure 4.1 illustrates the workflow of the two-step experimental design.

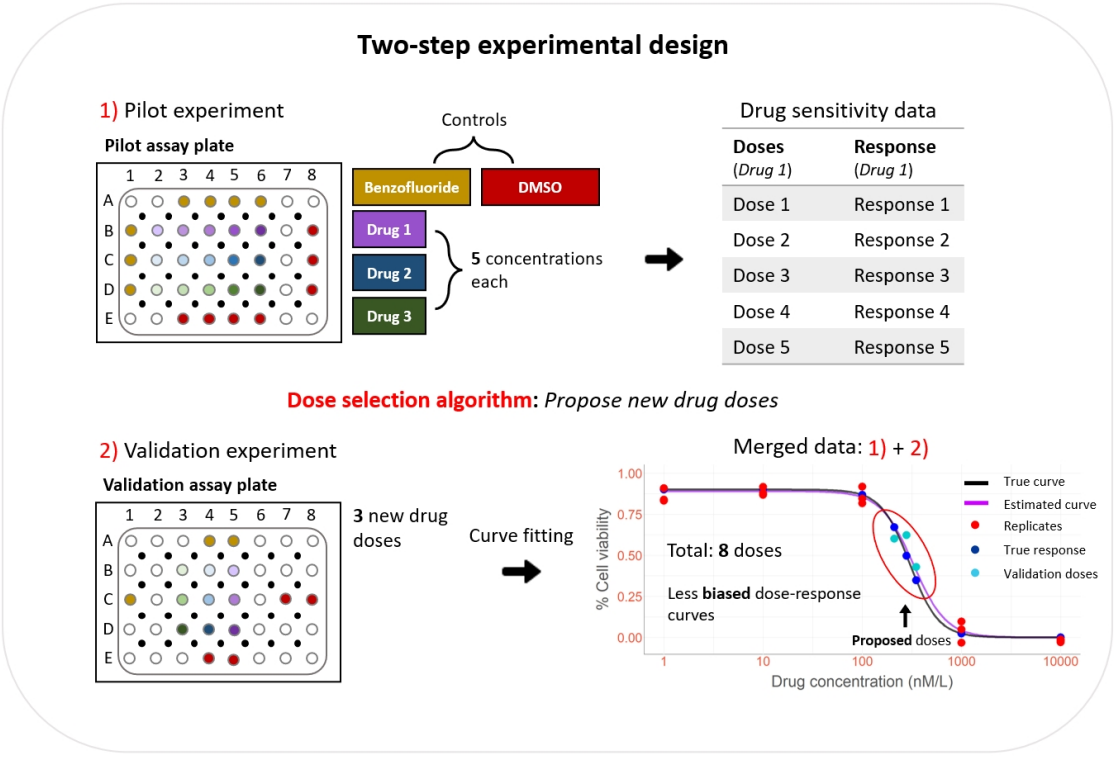

Figure 4.1: A two-step experimental design with the aim to reduce biases in dose-response estimation in the curve fitting.

4.2 A simulation workflow

4.2.1 The importance of computational simulations

Currently, there are mainly two branches of computational biology: high-scale simulations and knowledge discovery (Kitano 2002). Simulations are powerful methods to explore the dynamics of a biological system. In this thesis, simulations are applied to facilitate an understanding of the biases of different designs under a large variety of scenarios. In the context of pharmacological research and biomarker discovery, reducing DR curve fitting bias is crucial to accurately determine the potency of a drug. A related field that takes advantage of computational tools that aid biological experiments is in computational pharmaceutics (Ouyang et al. 2015). This field has been revolutionized by a harmonic relationship between computational molecular modelling with experimental research (Bunker et al. 2020).

4.2.2 Simulating drug sensitivity data

In order to validate the two-step experimental design via simulations, sampled dose-responses that reflect reality (i.e., viability values from actual experiments) is required. It is well-documented that bias in experimental procedures such as HTS is the result of both technical and biological variations (Bushway et al. 2010; Larsson et al. 2020; Niepel et al. 2019). Therefore, the *true* response for a specific cancer cell line in the presence of a compound is assumed randomly drawn from Normal distribution with mean μ centred at the true DR curve and with σ of 0.05.

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \tag{4.3}$$

A dose-response curve by sampling from the Equation 4.3 is illustrated in Figure 4.2.

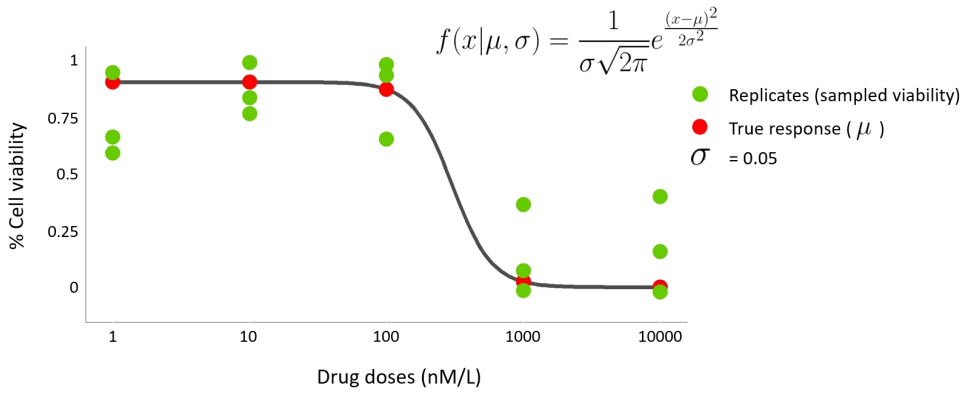

Figure 4.2: Simulation of a drug sensitivity experiment using normal distribution with μ centred at the true response and σ of 5%.

4.2.3 IC50 as a critical region in curve fitting

Exploratory computational analysis of DR curve fitting with **drc** indicates that the highest bias between the true and simulated DR curve is observed within the IC50 region (Figure 4.3). The workflow of this analysis is as such: (*1*) drug doses were computed in five different concentrations in a 10-fold format, (*2*) true parameters of the 4-parametric logistic function (Equation 2.1) were simulated, (*3*) three replicates were drawn from the Normal distribution (Equation 4.3) with the mean μ being the true response and a standard deviation σ of 0.05, (*4*) one replicate data fitted to the **drc** algorithm, (*5*) propose a new drug dose based on the DR curve, and finally (*6*) plotting the true and simulated DR curve alongside replicates, the true cell viability values, and the proposed drug dose. If the IC50 region is not properly covered by drug doses, **drc** is not able to properly estimate the true IC50 value which in turn affects current drug sensitivity quantification metrics and inaccurate inferences. However, adding a single

drug dose close to the correct IC50 region can significantly decrease the bias and therefore improve DR relationship inferences (Figure 4.3).

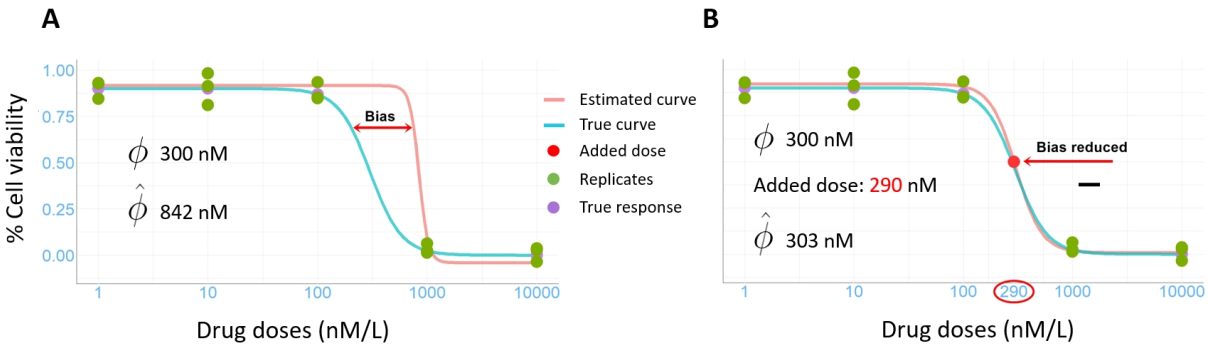

Figure 4.3: ϕ (IC50) as an essential parameter of the four-parameter logistic function. **A.** Bias identification within the ϕ region due to poor choice of drug doses. **B.** Adding a single drug doses close to the true ϕ value can considerably reduce the bias.

4.2.4 Generating dose-response curves

For the **drc** package, the 4-parameter logistic model *fct = L.4()* was selected along with the previously mentioned *BFGS* method, and the number of iterations was fixed at *10000*. In a simulation paradigm, the accuracy of the algorithm was computed based on the (Equations 4.1, 4.2) and determined both inference accuracy and the effectiveness of the two-step experimental design. In the experimental validation stage, the design was evaluated on the basis of comparing DR curves between high amounts of drug doses and biological replicates (i.e., *ground truth*) to the fitted curve proposed by the design.

4.2.5 Dose selection algorithm

Ideally, knowing the true IC50 region the solution would simply be to add drug doses within the IC50 region. However, for an actual experimental, we do not know the true

IC50 beforehand. Indeed, it is possible to infer the ϕ parameter using current R packages (Section 2.3), however, these are simply estimates and one wrong estimated parameter can ruin DR inference upon which using constantly such values to select drug doses is a naive approach. As such, a novel strategy to select efficient drug doses is crucial and required to tackle reproducibility by reducing inference bias in the curve fitting step.

The assumption of experimental bias in drug sensitivity experiments acknowledged in this thesis work is approximately 5% based on the observed variation between replicates from public drug sensitivity datasets. Assuming that each experimental replicate can be derived from a Normal distribution (Equation 4.3) and taking into account the critical region for the IC50 of *0.5* (% cell viability for IC50 concentration), along with standard deviation of 0.05, level of 2 (degree of deviation from the mean), and simulating 100000 cell responses, it was identified the range of *[0.4, . . . , 0.6]* to develop further the dose selection algorithm. This ultimately indicates that most of the viability values around 0.5 would fall in the reported range due to bias, landing in the 95% confidence interval landscape (Figure 4.4).

The approach to select new drug doses is based on four key settings: (*1*) cell viability values from the initial pilot experiment or previously reported results in respect to its dose concentration range, (*2*) minimum and maximum validation testing doses to be considered, (*3*) how much bias or errors in the experimental data is expected, and (*4*) level (i.e., how much deviation from the critical region of 0.5 is assumed) based on a Normal distribution paradigm - the higher the level, the farthest the viability values corresponding to 0.5 will deviate. Then, the algorithm suggests which and how many drug doses to add in the second round of a drug sensitivity experiment (i.e., *validation*). Finally, drug sensitivity data from both pilot and validation experiments are merged and a curve is fitted which is expected to represent the unbiased cellular response.

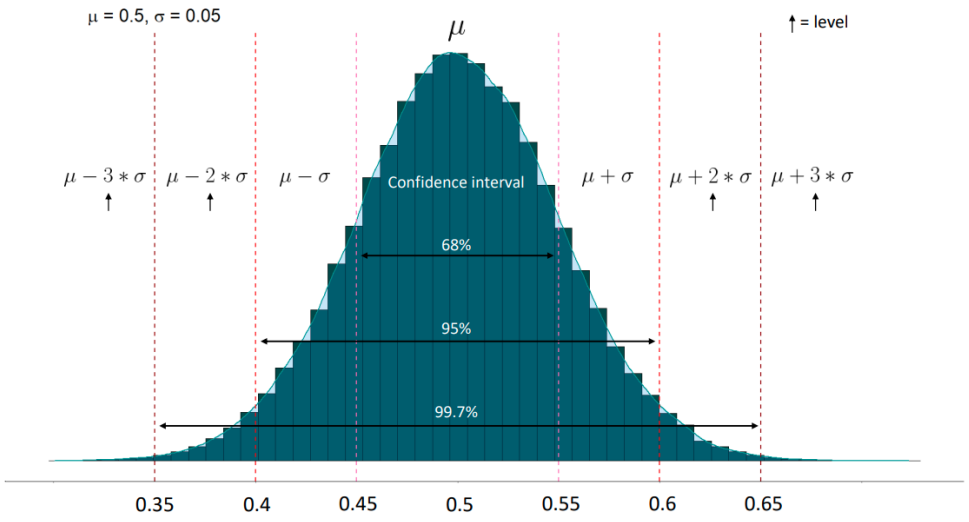

Figure 4.4: Normally distributed simulated cell viabilities. Critical region detected to select viabilities from [0.4 to 0.6] with μ of 0.5, σ of 0.05, and level of 2.

More formally, the algorithm works in a cascade of conditions that need to be met to choose new doses based on the *four* main cases. For instance, if the first condition is already satisfied, new doses will be determined. If not, the algorithm goes down the cascade until a specific condition is met. Once the algorithm is initialized, if the IC50 is already reached with the first drug dose (Figure 4.5, case **1**, highly effective compound within the tested dose range), then new doses are proposed prior to the initial pilot dose to cover the critical region of the DR curve. On the contrary, if the IC50 is not reached with the last dose from the pilot experiment (Figure 4.5, case **2**, ineffective compound within the tested dose range), drug doses beyond the last pilot dose will be added as an attempt to reach the IC50 region. Case **3** is a special case upon which all the viability values fall between the computed range of [0.4, . . . , 0.6] (ideally, without any bias, the viability for IC50 concentration would be simply 0.5). In such a paradigm, new drug doses are added before and after the first and last dose from the pilot experiment, respectively, making the dose interval wider. Furthermore, the majority of drug

sensitivity data will fall in the very last condition (case **4**) when all the previous conditions were not true. In such a case, the algorithm begins by finding the *first* dose followed by the *last* dose to cover the critical regions of the DR curve.

First and last dose determination

To find the first dose, the algorithm uncovers pilot concentrations whose viabilities are either above or below the threshold of 0.6 (i.e., $\mu = 0.5$, $\sigma = 0.05$, level 2). If the first cell viability is below 0.6, it means that the first dose from pilot is likely not enough and, thus, the minimal testing dose should be decreased in the validation step to ($Dmin$) (Figure 4.5, case **4**, first dose, **A**). Secondly, from all the viability values, if only the first viability is above 0.6, its respective concentration is considered as the first for validation (Figure 4.5, **case 4**, first dose, **B**). If all the viability values are consistently above 0.6 prior to the one that is below 0.6, the last dose whose viability is above 0.6 is considered for additional testing (Figure 4.5, **case 4**, first dose, **C**). If all of the above conditions are not met, the algorithm catches when the first viability is not above 0.6 and chooses the previous dose (Figure 4.5, **case 4**, first dose, **D**). To find the last dose, the algorithm uncovers pilot concentrations whose viabilities are either above or below the threshold of 0.4 (i.e., $\mu = 0.5$, $\sigma = 0.05$, level 2). In an event that the last viability value is above 0.4, the validation maximal testing dose will be increased to ($Dmax$) (Figure 4.5, **case 4**, last dose, **A**). Secondly, from all the viability values, if only the last is below 0.4, its respective concentration is considered as the last dose for validation (Figure 4.5, **case 4**, last dose, **B**). If all the viability values are consistently below 0.4 after the one that is above 0.6, the first dose whose viability is below 0.4 is considered for additional testing (Figure 4.5, **case 4**, last dose, **C**). If all of the above conditions are not met, the algorithm catches when the last viability is above 0.6 and chooses the next dose (Figure 4.5, **case 4**, last dose, **D**). Finally, doses are computed and proposed uniformly within the optimized range, to be tested for the validation experiment.

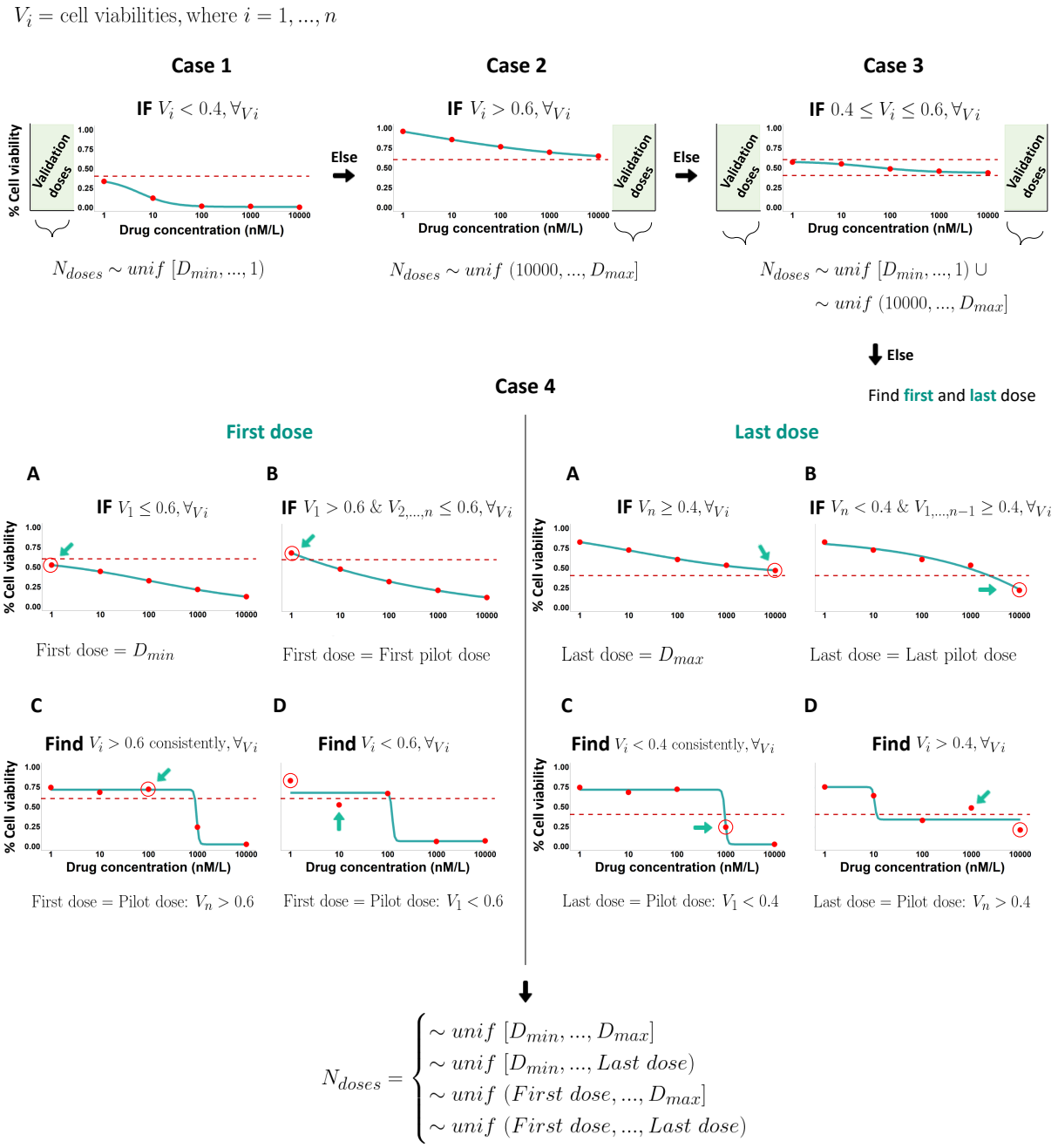

Figure 4.5: Algorithm to effectively select drug doses that cover the critical regions of the dose-response curve. *Ndoses* is the desired amount of doses for the validation experiment, *Dmin* is the minimal testing dose and *Dmax* the maximal testing dose. On each plot, **x** axis represents the drug concentration range (in log10 scale) and **y** axis represents the cell viabilities. Each curve was fitted using the **drc** R package under the four-parameter logistic model 2.1.

4.2.6 Parameter settings

To assess whether the proposed two-step experimental design by efficient dose selection could indeed help reduce bias in DR curve fitting, a massive series of simulations were performed with the **drc** R package. The 4-parameter logistic model (Equation 2.1) was used and its parameters were fixed based on the exploration of the parameter effects on a dose-response curve (Section 2.3.2) where a total of 4 β and α curve bounds were considered to cover distinct response scenarios, *41* values of η from -20 to -0.1 with a step of 0.5 to assess design performance varying the steepness of the curve, and finally *60* ϕ values on the grounds that ϕ affects highly DR inference based on the dose interval (Section 4.3) and, thus, ϕ values before, between, and after the tested doses were of interest. Further, 3 different number of replicates to simulate biological or technical replicates was considered as well as 3 different numbers of validation doses to evaluate design performance. In addition, was chosen a standard deviation (σ) of *0.05* and level of *2* highlighting how much deviation from the mean μ (*0.5*) is expected from the Normal distribution (Equation 4.3, and Figure 4.4). For the pilot drug doses, in total, 3 different numbers of doses were computed based on two forms of concentration intervals commonly used in pharmacological tests (i.e., log 10-fold and uniform intervals). Finally, for each parameter set, *100* simulations were performed. The code was developed to run at high efficiency where η and ϕ were set to run in parallel for 100 simulations each combination of the two for a specific parameter set. In total, *864* parameter sets comprising a total of *212544000* ($864 \times 41\eta \times 60\phi \times 100$ simulations) rows of data were computed and preprocessed to *2125440* ($864 \times 41\eta \times 60\phi \times 1$) where only summarized results out of 100 simulations for each set was of interest for downstream analysis.

Parameter	Value
Number of simulations	100
α	[0.0001, 0.1, 0.2, 0.3]
β	[0.7, 0.8, 0.9, 1]
η	[-20, ..., -0.1]
ϕ	[0.001, ..., 20000]
σ	0.05
Level	2
Dmin	0.0001
Dmax	25000
Number of replicates	[1, 3, 5]
Number of drug doses	[5, 10, 15]
Number of validation doses	[3, 4, 5]
Dose interval	[log 10-fold, uniform]

Table 4.1: Simulation parameters of the study.

4.3 Drug sensitivity experiments

4.3.1 Experimental protocol

For the experimental validation, drug sensitivity experiments were performed on *OPM-2* multiple myeloma cell lines. *OPM-2* cell lines were purchased from DMSZ (German Collection of Microorganisms and Cell Cultures GmbH) and routinely cultured in RPMI 1640 medium (Gibco/Thermo Fisher Scientific, Waltham, MA, USA) supplemented with

GlutaMax (Gibco), 10% fetal bovin serum, and antibiotics. Prior to the compound treatment, *OPM-2* cells were titrated to determine the optimal seeding number range for achieving a linear cell viability readout. Furthermore, around 2500 *OPM-2* cells per well were seeded in $25\mu L$ of culture medium in a 384-well plate (Corning no.3764, Corning, NY, USA), using an automatic dispenser BioTek MultiFlo FX RAD (5 μL cassette) (Biotek, Winooski, VT, USA). Selected drugs (see Table 4.2) at designed concentrations were subsequently added to the desired wells using a Tecan D300e digital dispenser (Tecan, Männedorf, Switzerland). For the positive control (toxic that is expected to kill all the cells), 100 μM of benzethonium chloride (BzCl) was used and 0.2% of DMSO for the negative control (i.e., vehicle with no effect on cells). After 72h incubation at 37°C and 5% CO2, the plates were equilibrated at room temperature for 30min before starting the to viability assay using 25 μL per well CellTiter-Glo luminescent reagent (Promega, Madison, WI, USA). Finally, luminescence signal was recorded after 10 min incubation with a PHERAstar FS multimode plate reader (BMG Labtech, Ortenberg, Germany).

Drug	**g/mol**	**Supplier**	**Target**	**Class**
Ruxolitinib	310.00	ChemieTek	JAK1/2 inhibitors	Kinase inhibitor
Axitinib	386.47	LC Laboratories	VEGFR, PDGFR, KIT inhibitor	Kinase inhibitor
Carfilzomib	719.91	ChemieTek	Proteasome inhibitor (20S subunit)	Protease inhibitor
Bortezomib	384.24	ChemieTek	Proteasome inhibitor (26S subunit)	Protease inhibitor
Vorinostat	264.32	LC Laboratories	HDAC inhibitor	Kinase inhibitor
AZD4547	463.58	ChemieTek	FGFR inhibitor	Epigenetic modifier
Lonafarnib	471.35	Selleck	PDK1 inhibitor	Kinase inhibitor

Table 4.2: A list of drugs used in the experimental validation assays.

Chapter 5

Results

5.1 Assessment of the two-step design in simulated dose-response data

5.1.1 Adding replicates does not reduce the bias

In the majority of drug sensitivity experiments, researchers tend to include multiple technical or biological replicates to improve the estimation of the dose-response curves for a particular drug (Bolomsky et al. 2021). An overview of the selective true DR curves with the chosen parameters depicted in the plot, suggests that depending on the ϕ, the dose-response curve will behave differently (Figure 5.1, **A**). However, supported by simulations, it was found that adding replicates does not necessarily reduce the DR inference bias across several ϕ values (Figure 5.1, **B**, **C**, and **D**). This is the result of poor ϕ coverage by the tested doses. Such conclusion can be made for every single parameter set which suggests the need to develop computational models to (*1*) identify the critical regions of the DR curve, and (*2*) propose new doses to improve DR

estimation rather than the addition of replicates.

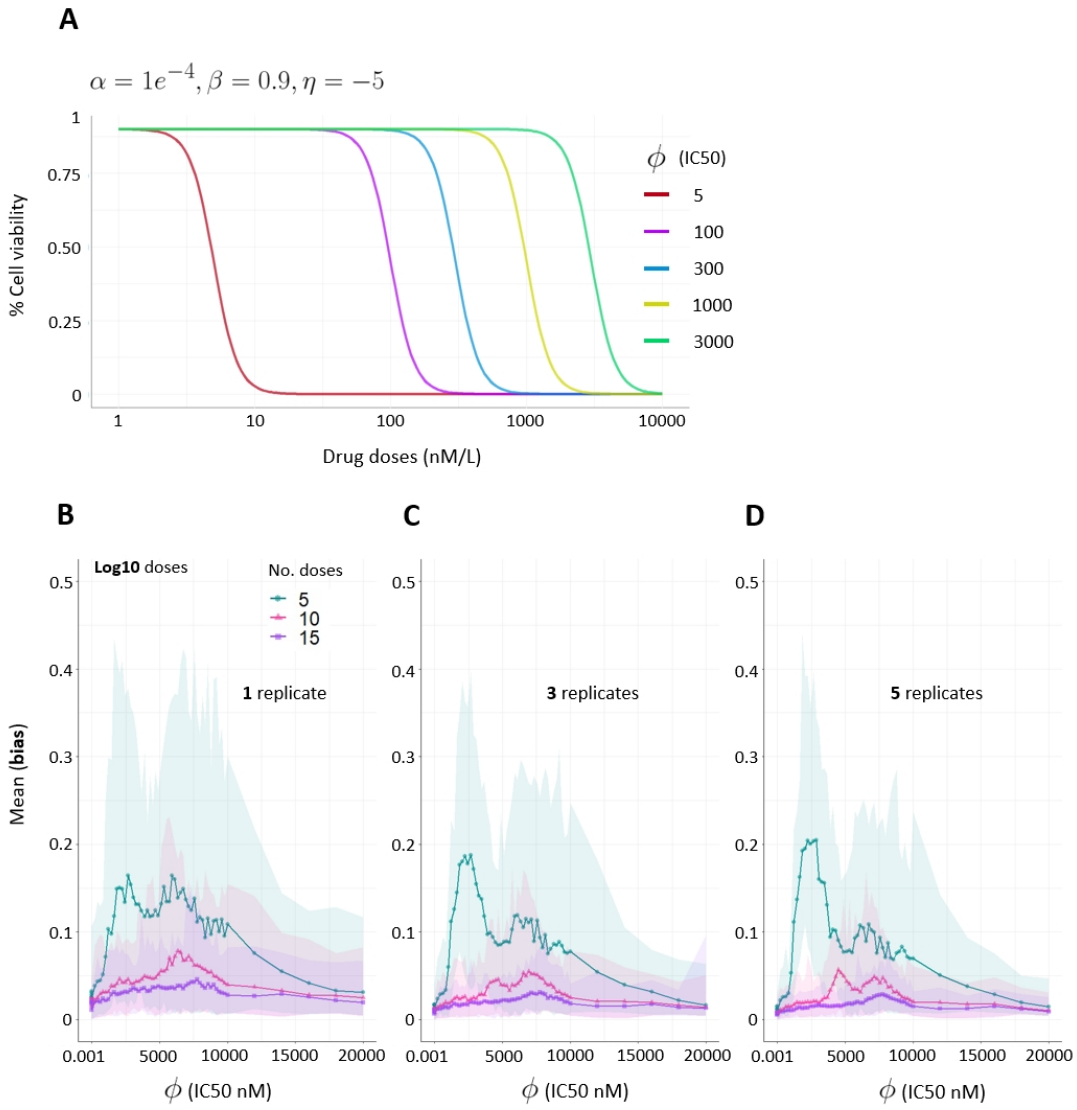

Figure 5.1: Adding replicates does not improve the estimation of dose-response curves. **A**. Overview of the true dose-response curves for fixed α, β, and η parameters, varying ϕ. **B**. Bottom row plots illustrate the absolute mean of bias (Equation 4.1) as an average over 100 simulations across 60 different ϕ values for a specific parameter set of the logistic model. The plot highlights absolute mean of bias for pilot experiment in a log10 interval with 1 replicate, **C**. bias for 3 replicates, and **D**. bias for 5 replicates.

5.1.2 Bias is significantly reduced following the new dose proposal by the two-step design

To evaluate the performance of the dose selection algorithm through the two-step design workflow, the absolute mean of bias of 100 DR curves to the ground truth for each of the parameter combinations (i.e., *864*) was evaluated. For each parameter set, 100 DR curves are estimated for (η - ϕ) combination and the average performance is determined. Three key metrics were used to assess the performance of the design compared to a single experiment (i.e., pilot) with replicates: (**1**) absolute mean of bias (Figure 5.2), (**2**) mean relative error (RE) of AUC (Figure 5.3), and (**3**) median RE of ϕ (Figure 5.4) in log10 scale due to high variation within the values. The reason for using median instead of mean for ϕ is that the variance of the estimates increased drastically for some IC50s (ϕ) which makes the mean estimation less robust. It is observed that poor inference is present when the ϕ region is not properly covered by tested doses (i.e., pilot plots on each Figures 5.2, 5.3, and 5.4). Followed by a second round of simulated experiments with newly proposed doses and merging data from both pilot and validation at different amounts, the inferred DR curves across every ϕ are indistinguishable to the ground truth due to low bias and relative errors, *suggesting improved performance of the two-step design*. In addition, a key observation is that, by simply adding 3 validation doses to the initial 5 from the pilot experiment, the inference accuracy is superior compared to using 15 doses and 5 replicates (i.e., 8 cell culture wells as opposed to 75), thereby optimizing cell viability assays. It was found that the addition of 3 or 5 validation doses with one replicate provided similar results, suggesting that the inclusion of 3 validation doses to the pilot experiment is sufficient to infer accurate DR curves. Moreover, the addition of 5 doses for absolute mean of bias attenuated bias for IC50s between 1000 and 5000, which is expected since more doses will cover this critical

region. For mean RE of AUC and median RE of ϕ, the two-step design decreased wrongly estimated DR curves across 100 simulations (shades) for every number of pilot drug dose, suggesting superior confidence while estimating drug effects with such drug sensitivity metrics. As ϕ increases, the more challenging is to lower relative errors, still the two-step design diminished both AUC and ϕ. Comparing doses in different intervals, both the absolute mean of bias and mean RE AUC is lower when doses are in uniform scale, due to proper IC50 coverage as opposed to ϕ parameter estimation (depending on how the algorithm chooses the best parameter values in the curve fitting). Taken together, the proposed two-step design outperforms a single experiment for every drug sensitivity metric across all IC50 (ϕ) values at both different dose concentration intervals and amounts through the use of simulations.

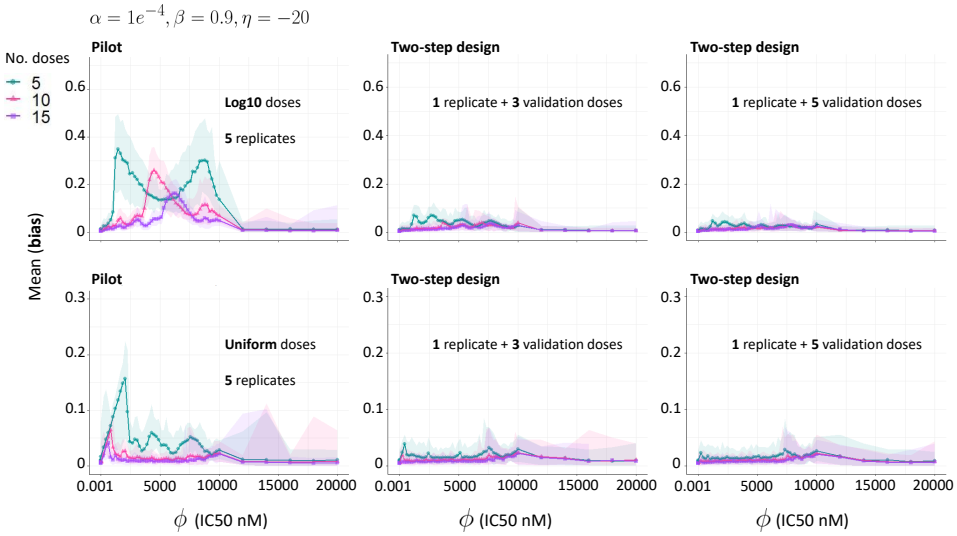

Figure 5.2: Two-step design performance for absolute mean of bias for different number of pilot doses across 60 ϕ values. Model parameters are highlighted at the top of the plot. First row highlights doses in log10 and bottom row in uniform interval. Shaded areas spans between the minimum and maximum bias values.

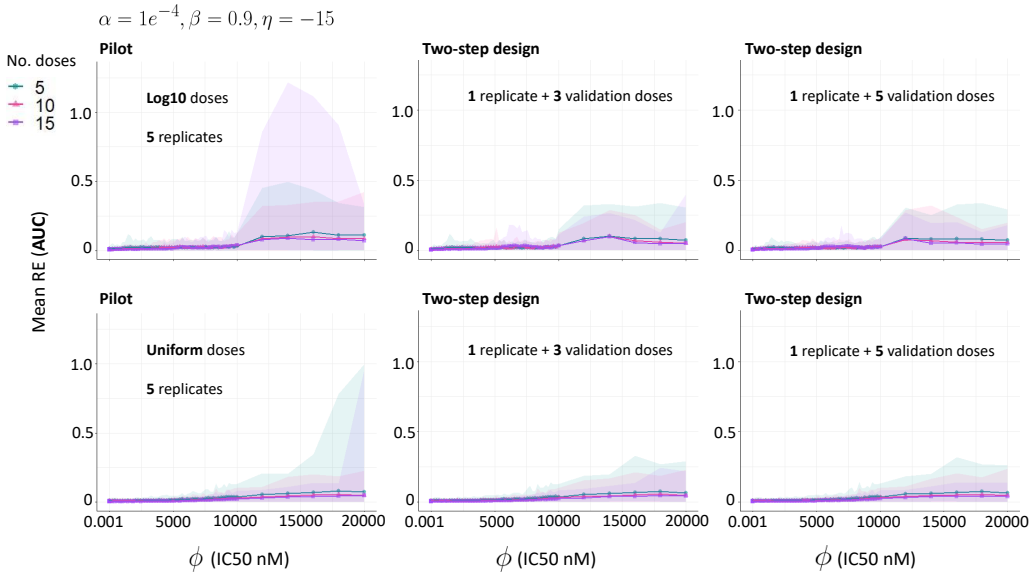

Figure 5.3: Two-step design performance for mean RE of AUC for different number of pilot doses across 60 ϕ values. Model parameters are highlighted at the top of the plot. First row highlights doses in log10 and bottom row in uniform interval. Shaded areas spans between the minimum and maximum RE AUC values.

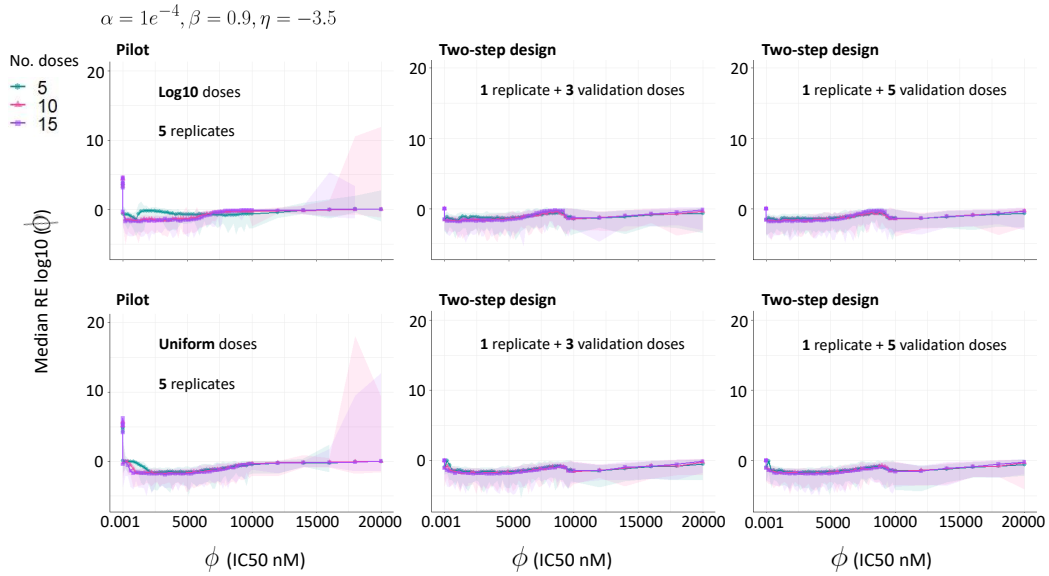

Figure 5.4: Two-step design performance for median RE of ϕ for different number of pilot doses across 60 ϕ values. Model parameters are highlighted at the top of the plot. First row highlights doses in log10 and bottom row in uniform interval. Shaded areas spans between the minimum and maximum RE ϕ values.

5.1.3 Bias is minimized once slope goes to zero

A closer look into the distribution for absolute mean of bias for similar parameters from previously reported analysis, suggests that when η tends to zero, the two-step design is able to infer proper DR curves with the addition of 3 validation doses from pilot compared to 15 (Figure 5.5). On both boxplot distribution plots, absolute mean of bias was measured across all ϕ values for a specific parameter set of the logistic model for doses in log10 and uniform intervals. As expected, regarding pilot experiments, using 15 drug doses consistently reduces bias compared to 5 for both intervals. With the addition of 3 validation doses for the two-step design, the bias is significantly decreased compared to a single experiment with 15 doses and 5 replicates when slope goes to 0. In total, bias

was tested for all of the 60 IC50s (ϕ). Outcomes of such distribution is generalized for every parameter combination of the study, suggesting good performance of the two-step design with the addition of 3 doses to the initial 5 from the pilot experiment.

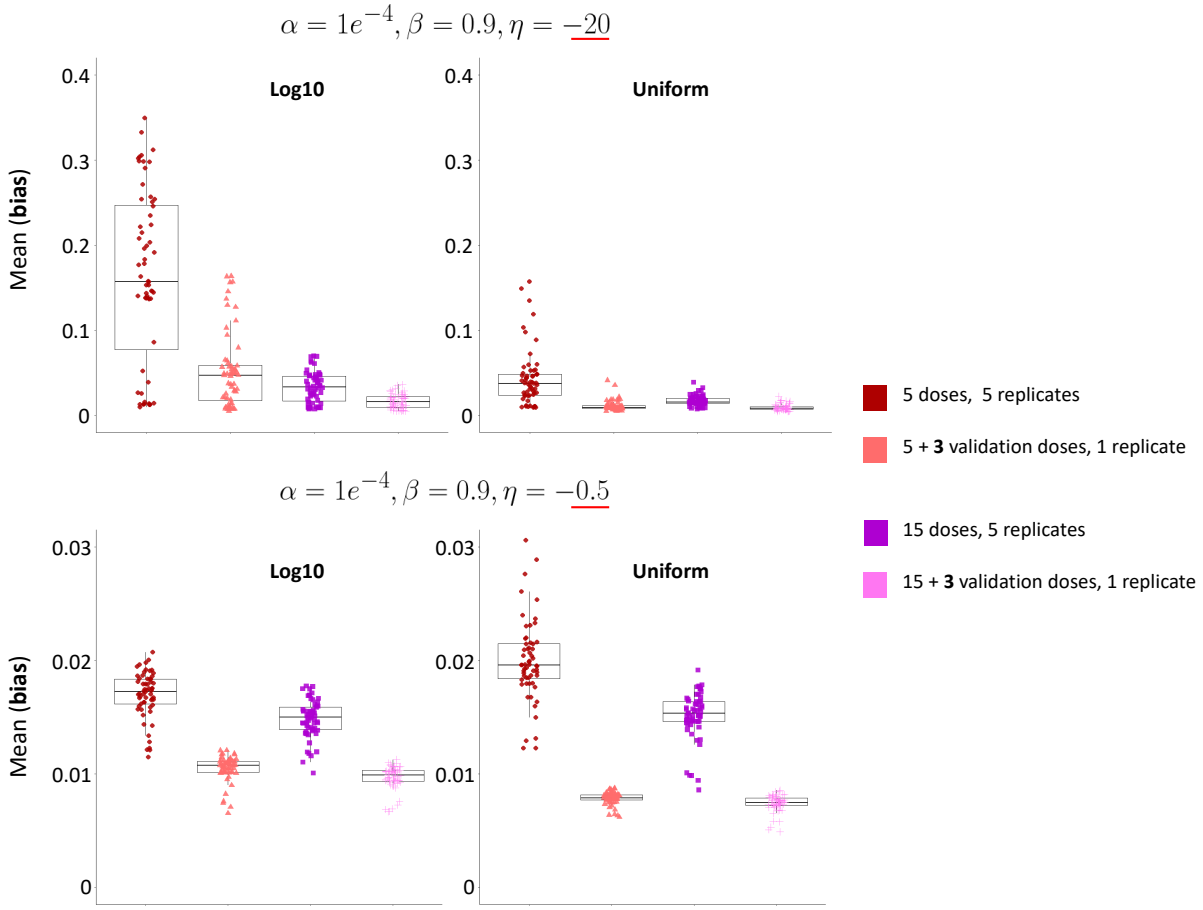

Figure 5.5: Boxplots highlighting bias distribution for different number of doses, concentration range, and ϕ.

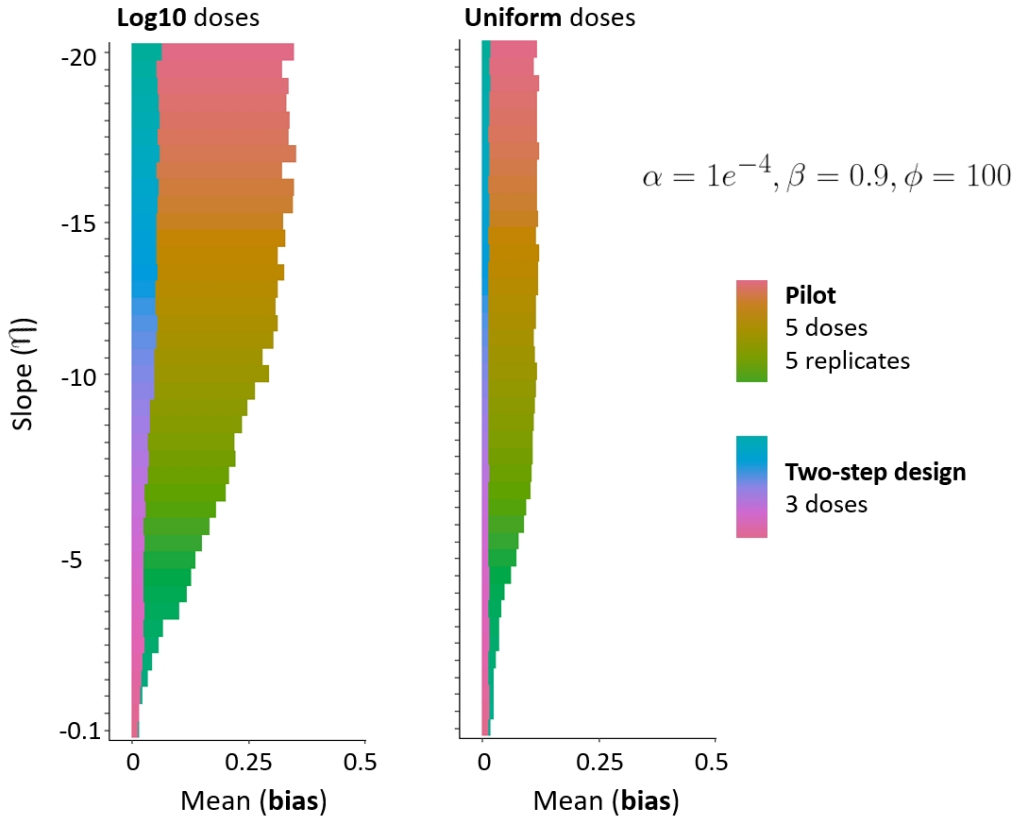

Figure 5.6: Relationship between absolute mean of bias and η for both the pilot and the two-step design, at fixed ϕ of 100.

As a result of observing differences regarding absolute mean of bias for the lowest and highest η value, further analysis was conducted to detect the relationship when $-20 \leq \eta \leq -0.1$ (Figure 5.6). For similar parameters, it is indeed easier to infer accurately DR curves when the steepness of the *true* curve goes towards zero. However, in practice, this zero steepness is uncommon since drugs and cells usually display a sigmoidal relationship (Malyutina et al. 2021).

5.1.4 Wide exploratory statistics

Next, a full-scale statistical analysis for all the 864 parameters of the simulation workflow was conducted for the three key metrics (bias, AUC, ϕ). First, an overall distribution for the absolute mean of bias for both the pilot and the two-step design was computed. Analogous to the previous findings, the design improved DR inference in such a way that bias did not go above *0.1* whilst bias from the pilot experiments can reach up to *0.4* with much higher frequencies (Figure 5.7).

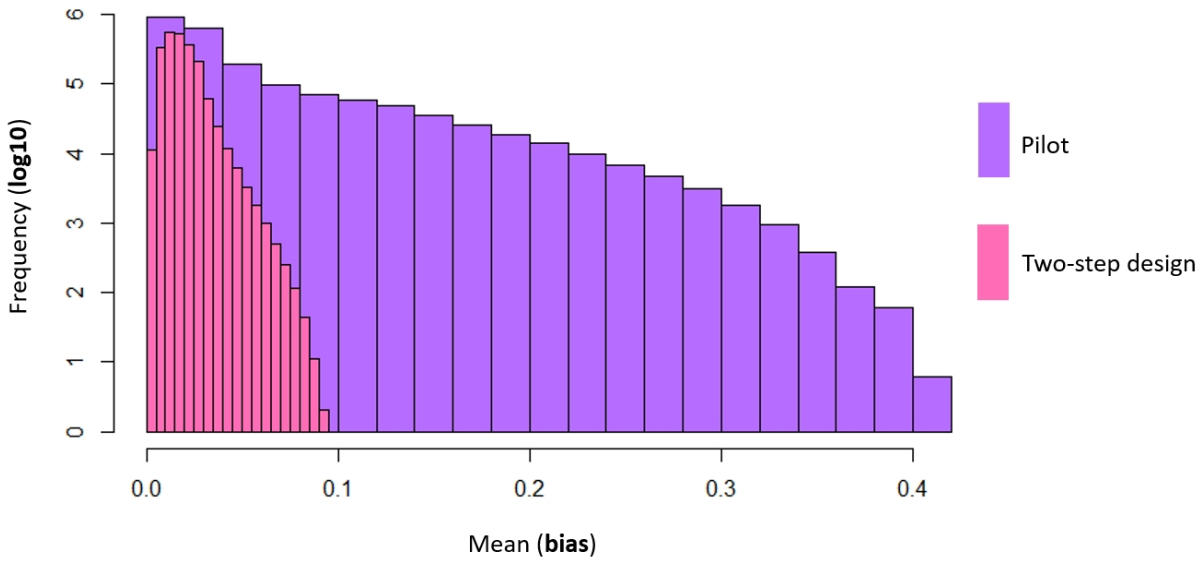

Figure 5.7: Histogram for mean of bias across all the parameter sets for both the pilot and the two-step design.

Finally, summary statistics for (*1*) absolute mean of bias, (*2*) mean RE of AUC, and (*3*) median RE of ϕ emphasizes the power of drug dose selection and how highly similar estimated DR curves are compared with a single experiment with many replicates for the two-step design (Table 5.1). Regarding bias, and as expected, the design

outperformed pilot across every scenario whilst attenuating the maximum possible bias across 100 simulations for every parameter set. For AUC, the design outperformed pilot for initial doses in log10 format. However, whenever initial doses were uniformly distributed, since the theoretical expectation of ϕ was covered in a more consistent manner, it was difficult to beat the pilot experiment (although the results are quite similar). Regarding ϕ, mean across all median RE of ϕ highlights how accurate the design is able to correct ϕ estimates (even more accentuated by comparing maximum estimated values), aiding in properly estimating the drug effect on cells since the ϕ parameter of the logistic model is widely used and recommended across the scientific community to detect drug potency (Larsson et al. 2020).

		Pilot					**Design**				
			Bias					Bias			
Dose no.		Min	Mean	Median	Quantiles	Max	Min	Mean	Median	Quantiles	Max
5	Log10	0.007	0.101	0.093	(0.03, 0.14)	0.414	0.006	**0.024**	0.024	(0.01, 0.02)	0.09
	Uniform	0.007	0.033	0.027	(0.02, 0.04)	0.177	0.006	**0.018**	0.017	(0.01, 0.02)	0.05
10	Log10	0.005	0.05	0.037	(0.02, 0.06)	0.286	0.004	**0.018**	0.018	(0.01, 0.02)	0.096
	Uniform	0.005	0.017	0.015	(0.01, 0.02)	0.079	0.004	**0.013**	0.012	(0.01, 0.017)	0.033
15	Log10	0.004	0.032	0.024	(0.01, 0.03)	0.18	0.003	**0.016**	0.016	(0.01, 0.02)	0.070
	Uniform	0.004	0.013	0.011	(0.009, 0.01)	0.054	0.003	**0.011**	0.01	(0.008, 0.01)	0.029
			AUC					AUC			
		Min	Mean	Median	Quantiles	Max	Min	Mean	Median	Quantiles	Max
5	Log10	0.01	20.98	0.03	(0.02, 0.05)	92421	0.01	**13.24**	0.03	(0.02, 0.04)	57275
	Uniform	0.01	**50.28**	0.02	(0.02, 0.04)	214198	0.01	50.46	0.02	(0.02, 0.04)	220557
10	Log10	0.01	22.44	0.03	(0.02, 0.04)	86787	0.00	**13.85**	0.02	(0.02, 0.04)	63644
	Uniform	0.00	**64.69**	0.02	(0.01, 0.03)	282367	0.00	66.96	0.02	(0.01, 0.03)	308472
15	Log10	0.00	25.83	0.02	(0.02, 0.04)	117147	0.00	**15.52**	0.02	(0.01, 0.03)	65011
	Uniform	0.00	**63.24**	0.01	(0.01, 0.02)	288881	0.00	64.74	0.01	(0.01, 0.02)	294105
			ϕ					ϕ			
		Min	Mean	Median	Quantiles	Max	Min	Mean	Median	Quantiles	Max
5	Log10	0.03	1789	0.54	(0.28, 1.00)	279535	0.005	**0.648**	0.068	(0.03, 0.22)	**3471**
	Uniform	0.00	38477	0.00	(0.00, 1.00)	7656105	0.003	**0.606**	0.038	(0.02, 0.24)	**5928**
10	Log10	0.00	1931	0.4	(0.1, 0.9)	415820	0.004	**0.623**	0.035	(0.01, 0.2)	**4041**
	Uniform	0.00	45444	0.00	(0.00, 1.00)	16250175	0.003	**0.581**	0.029	(0.01, 0.19)	**3568**
15	Log10	0.00	1982	0.3	(0.1, 0.8)	384690	0.003	**0.54**	0.03	(0.01, 0.22)	**3070**
	Uniform	0.00	49617	0.00	(0.00, 1.00)	14631926	0.002	**0.57**	0.02	(0.01, 0.18)	**6930**

Table 5.1: Summary statistics of performance for the two-step design compared to pilot.

5.2 Experimental validation

Given the complexity of the pilot assay plates, validation experiment analysis was conducted differently for each of the considered drugs. The workflow was as follows: (**1**) statistics for the pilot stage for each compound was collected and new drug doses were computed for validation (e.g., 3 and 5) based on the behaviour and theoretical expectation of the ϕ region using one replicate for both log10 and uniform concentration intervals, and (**2**) fit an estimated DR curve from both pilot and validation assays for comparison to the *ground truth* comprising a total of 15 doses and 3 replicates in a log10 scale. The following sections are focused on the outcomes of experimental validation for 7 drugs whereupon the two-step design outperformed a single experiment run for 5 drugs.

5.2.1 Vorinostat

Vorinostat is a well-known oral HDAC (histone deacetylase) inhibitor for the treatment of relapsed or refractory multiple myeloma (MM) patients (Siegel et al. 2014; Silva et al. 2013). Outcome of the *ground truth* suggests a sigmoidal relationship where half of the cells were inhibited within the dose region of [100, 1000] nm/L on Figure 5.8. DR data from the pilot experiments with one and five replicates in the uniform interval were not able to accurately infer the *true* behaviour, highlighting an estimation of IC50 sooner than the ground truth. On the contrary, the two-step design achieved consistent high accuracy with the estimated curve being in line with the *ground truth* (blue DR curves). To validate the findings for Vorinostat, ϕ values were compared for log interval whereupon the ground truth depicts an estimated $\hat{\phi}$ of 680 nM while, with a single pilot experiment, the estimated $\hat{\phi}$ was 573 nM and with the addition of validation doses of 648 nM (Figure 5.9). Thus, the estimated $\hat{\phi}$ from the two-step design is closer to the

ground truth, suggesting less biased DR inference. Regarding RE of AUC, the two-step design outperformed pilot with a value of 0.02 against 0.03. Finally, for RE of $\hat{\phi}$, and as expected, the estimation was close to the ground truth and, thus, a relative error of 0.04 against 0.16 suggests higher inference accuracy for the two-step design in scenarios where cells respond drastically whilst increasing drug dosage.

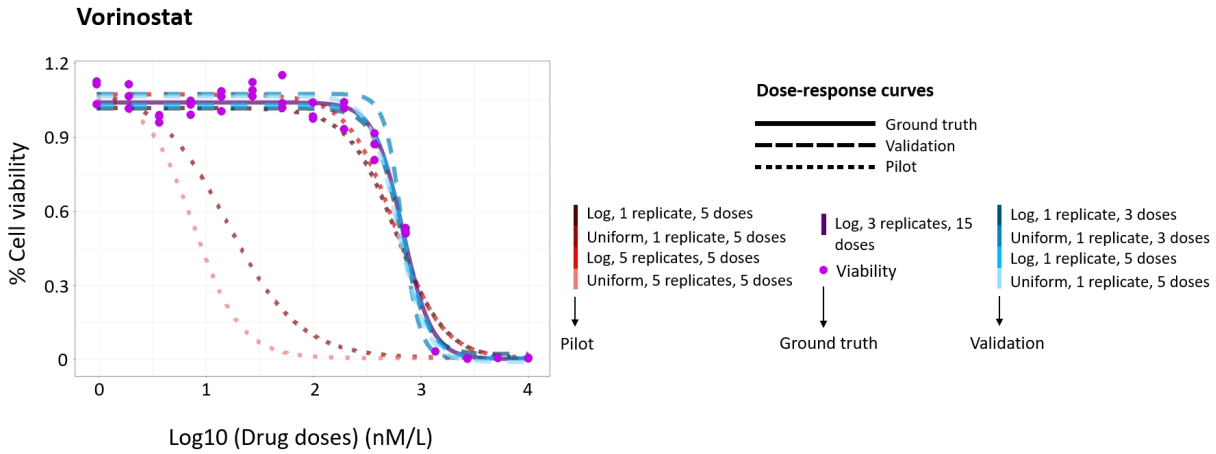

Figure 5.8: Validation assay analysis of design performance for Vorinostat.

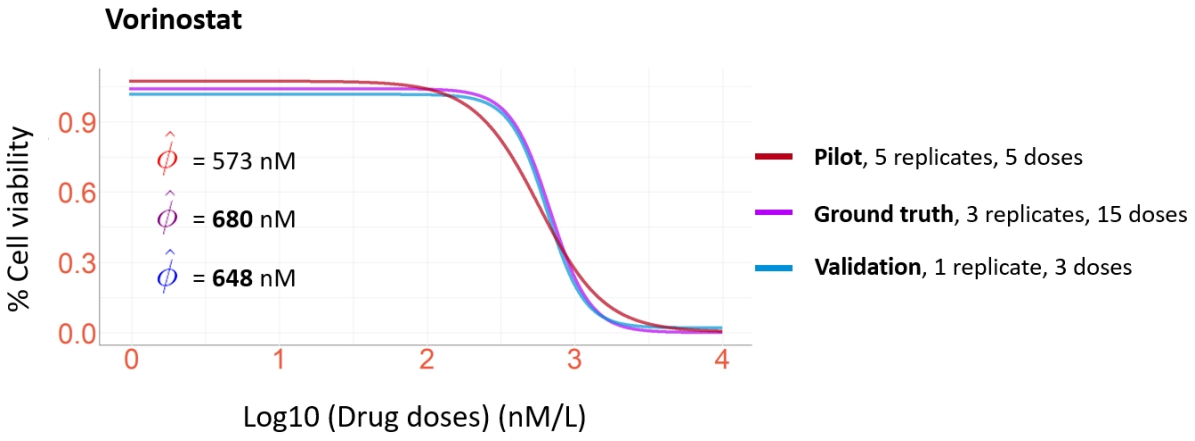

Figure 5.9: Comparison of dose-response curves determined by pilot, validation, and the ground truth on the log dose interval for Vorinostat.

5.2.2 AZD4547

AZD454, is a potent inhibitor of the firoblast growth factor (FGF) 1 and 2 by silencing GFG signalling and proliferation of tumor cells (Gavine et al. 2012). In comparison to Vorinostat, AZD4547 drug inhibits MM cells by 50% with smaller doses. It is observed that inferences from the pilot experiments will most likely result in non-harmonic outcomes since the steepness of these curves are not representative of the ground truth which is not the case for validation, highlighting the robustness of the two-step design for the AZD4547 drug (Figure 5.10). With validation doses, the DR curves generally display a sigmoidal relationship similar to the ground truth. To validate the findings for AZD454, ϕ values were compared for log interval upon which the ground truth depicts an estimated $\hat{\phi}$ of 161 nM while, with a single pilot experiment, the estimated $\hat{\phi}$ was 1218 nM and with the addition of validation doses of 333 nM (Figure 5.11). Thus, the estimated $\hat{\phi}$ from the two-step design is closer to the ground truth, suggesting less biased DR inference. As a result of poor ϕ coverage, the addition of replicates do not improve DR estimation in the curve fitting (effect validated through simulations) (Figure 5.1). Regarding RE of AUC, the relative error for the pilot experiment for doses in log interval was lower than the two-step design (0.002 against 0.01). Finally, for RE of $\hat{\phi}$, and as expected, the estimation was close to the ground truth and, thus, a relative error of 1 against 6.6 from the design suggests that in such DR scenarios, the estimation for the two-step experimental design is superior.

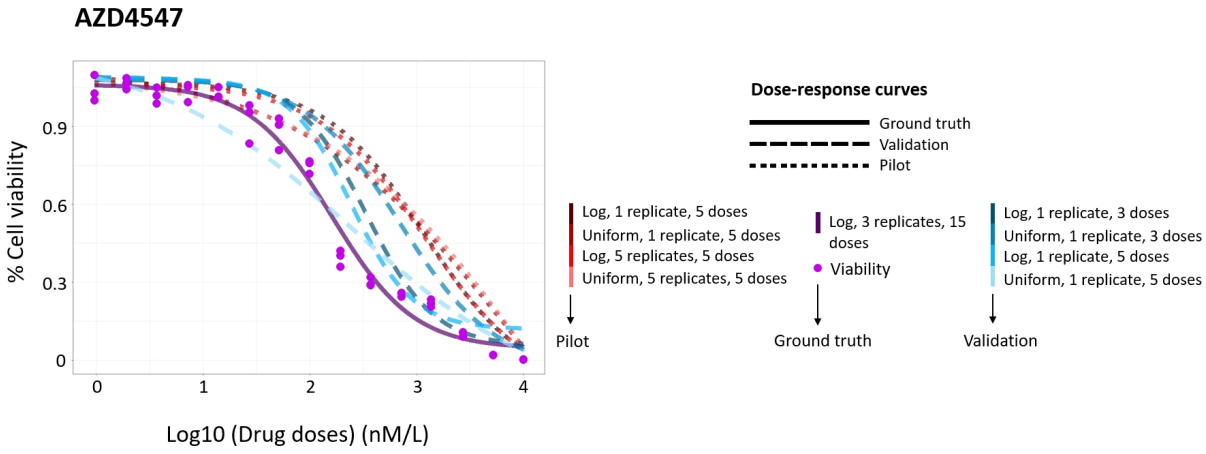

Figure 5.10: Validation assay analysis of design performance for AZD4547.

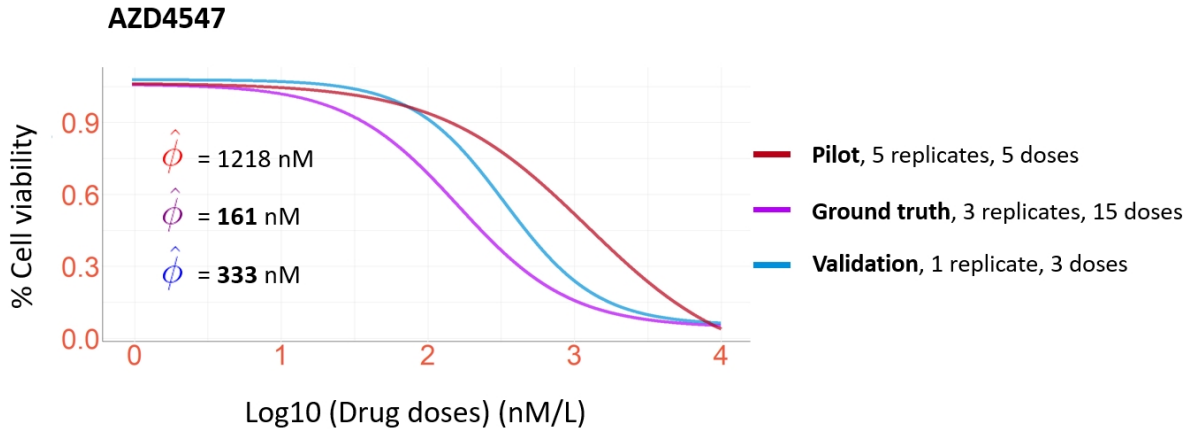

Figure 5.11: Comparison of dose-response curves determined by pilot, validation, and the ground truth on the log dose interval for AZD4547.

5.2.3 Axitinib

Axitinib is a frequently used drug to target mutant hematological cancer cells. Axitinib is a vascular endothelial growth factor (VEGF) receptor tyrosine kinase inhibitor with high affinity (McTigue et al. 2012). DR analysis of Axitinib on MM cell lines indicate that, first, is less potent than AZD4547 with a $\hat{\phi}$ of 1507 nM in respect to the ground

truth DR curve while AZD4547 the $\hat{\phi}$ value was around 161 nM (Figure 5.12). To validate the findings for Axitinib, ϕ values were compared for log interval upon which the ground truth depicts an estimated $\hat{\phi}$ of 1507 nM while, with a single pilot experiment, the estimated $\hat{\phi}$ was 1455 nM and with the addition of validation doses of 881 nM (Figure 5.13). Curiously, the cell viability assay results for this specific drug should indicate less biased DR curve from pilot due to the highly similar $\hat{\phi}$ regions. However, after curve fitting with **drc**, the two-step design curve is much more representative to a higher degree to the ground truth than the pilot curve (blue against red, respectively). This might be explained due to a better fitting from the other 3 parameters of the logistic model (α, β, and η). Regarding RE of AUC, the relative error for the pilot experiment for doses in log interval was higher than the two-step design by a margin of around 0.03 (0.07 to 0.04 from the design). Intuitively, for RE of $\hat{\phi}$, the pilot experiment estimated $\hat{\phi}$ was close to the ground truth and, thus, a relative error of 0.03 against 0.41 from the design was found.

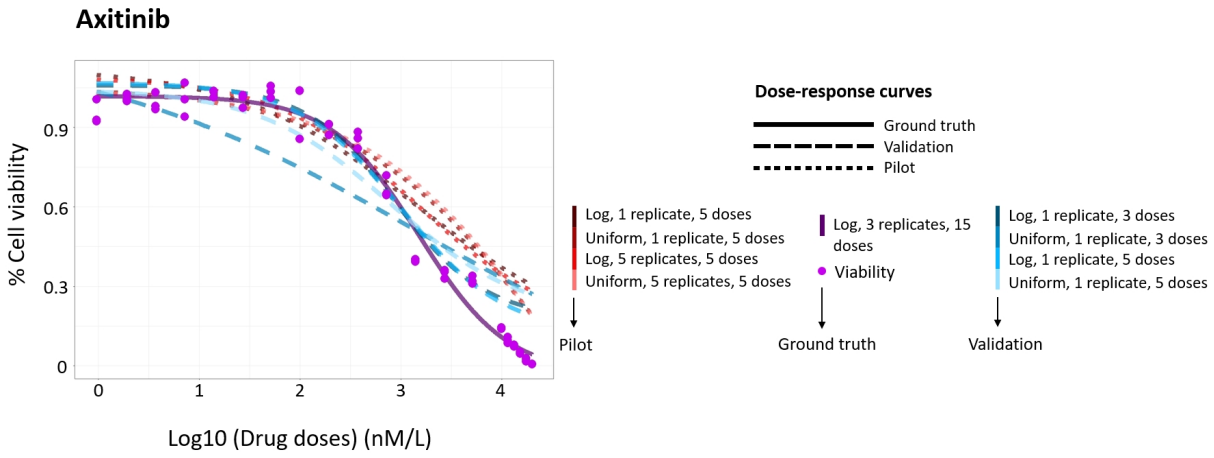

Figure 5.12: Validation assay analysis of design performance for Axitinib.

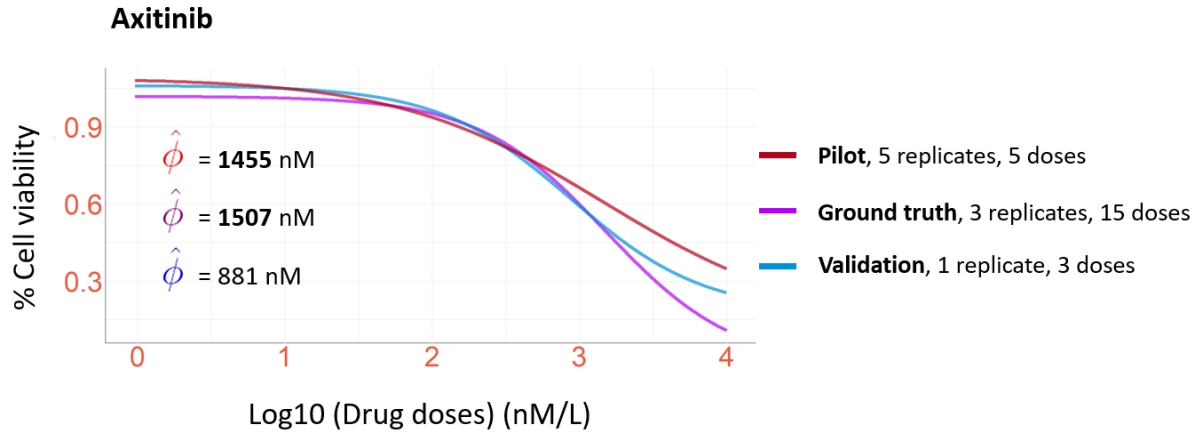

Figure 5.13: Comparison of dose-response curves determined by pilot, validation, and the ground truth on the log dose interval for Axitinib.

5.2.4 Ruxolitinib

Ruxolitinib is an inhibitor of JAK1/2 pathways which are essential for cell survival and proliferation (Mascarenhas et al. 2012). However, in MM cancer cells, Ruxolitinib is only able to inhibit MM cells at ϕ around 10000 nM (Figure 5.14) for doses in both intervals (log and uniform). To validate the findings for Ruxolitinib, ϕ values were compared for log interval upon which the ground truth depicts an estimated $\hat{\phi}$ of 17763 nM while, with a single pilot experiment, the estimated $\hat{\phi}$ was 2753 nM and with the addition of validation doses of 11516 nM (Figure 5.15). Albeit closer $\hat{\phi}$ to the ground truth for the two-step design, is clearly visible some inference errors provided by the algorithm during curve fitting. Nonetheless, it behaves in a similar fashion which is not the case for the pilot experiment with 5 replicates. Regarding both RE of AUC and $\hat{\phi}$, the relative errors for the two-step design were lower than for the pilot experiment (0.06 against 0.11 from pilot, and 0.35 against 0.84, respectively).

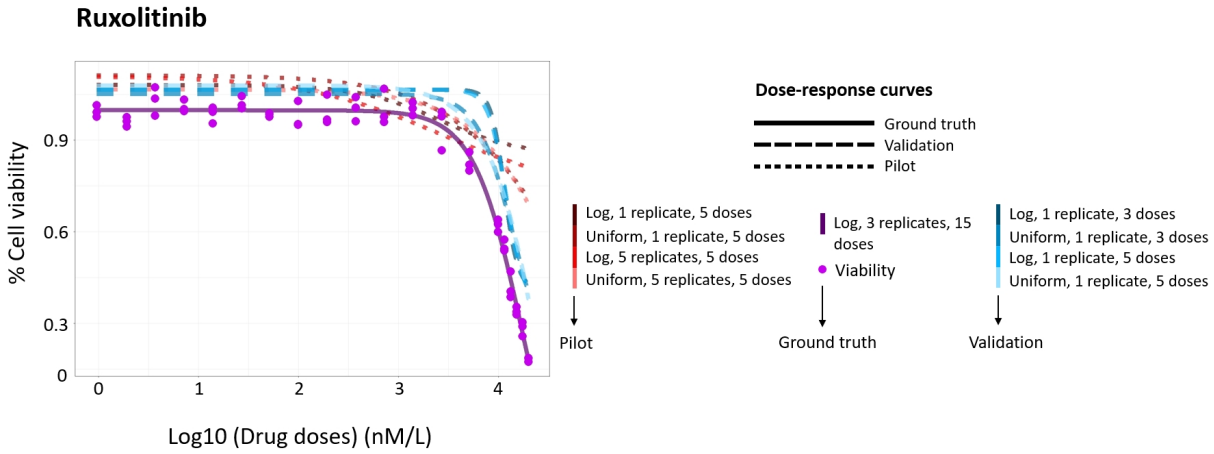

Figure 5.14: Validation assay analysis of design performance for Ruxolitinib.

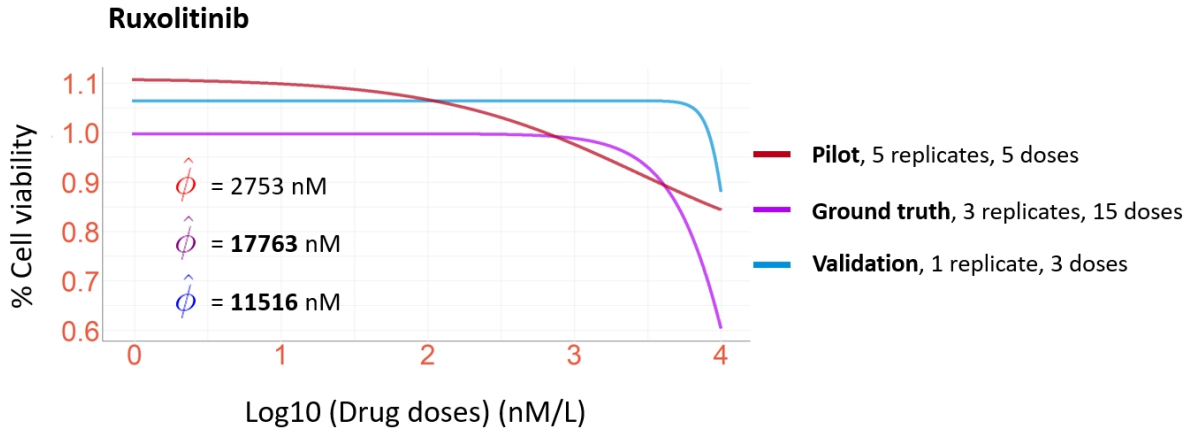

Figure 5.15: Comparison of dose-response curves determined by pilot, validation, and the ground truth on the log dose interval for Ruxolitinib.

5.2.5 Lonafarnib

Lonafarnib is a RAS signalling pathway inhibitor. RAS pathway has been reported to be responsible for tumor cell growth through RAS gene mutation effects (Downward 2003). Lonafarnib effects on MM cells are somewhat similar to Ruxolitinib in a way that is only able to inhibit half of cells with high concentrations within the range of

102024nM (i.e., *ground truth*) (Figure 5.16). DR inference in cases such as for Lonafarnib, the addition of replicates will not help since the tested doses from the pilot experiment will most likely not cover ϕ regions above its respective maximum tested dose (thus the relevance of the dose selection algorithm - Figure 4.5). To validate the findings for Lonafarnib, ϕ values were compared for log interval upon which the ground truth depicts an estimated $\hat{\phi}$ of 102024 nM while, with a single pilot experiment, the estimated $\hat{\phi}$ was 11038 nM and with the addition of validation doses of 11222 nM (Figure 5.17). Indeed, the two-step design with the addition of 3 validation doses as an attempt to reach the true ϕ makes the DR inference better than pilot with high amounts of replicates, however, the estimation still suffers which might be explain by both batch effects and drivers of variability in the screening assay (Section 2.2.2). Interest findings for both RE of AUC and $\hat{\phi}$ highlights how challenging it is to infer DR curves when the assumptions under the logistic model are not met. In respect to RE of AUC, the relative error for pilot was lower than the two-step design by a margin of 0.11 (0.05 to 0.16 from the design). Remarkably, RE of ϕ for the two-step design was better than pilot by a margin of 0.1 (0.89 to 0.88 from pilot) which can be explained by the fact that, even if the $\hat{\phi}$ from the design is closer to the ground truth than the estimation from the pilot experiment, it was still insufficient to overperform by a significant margin.

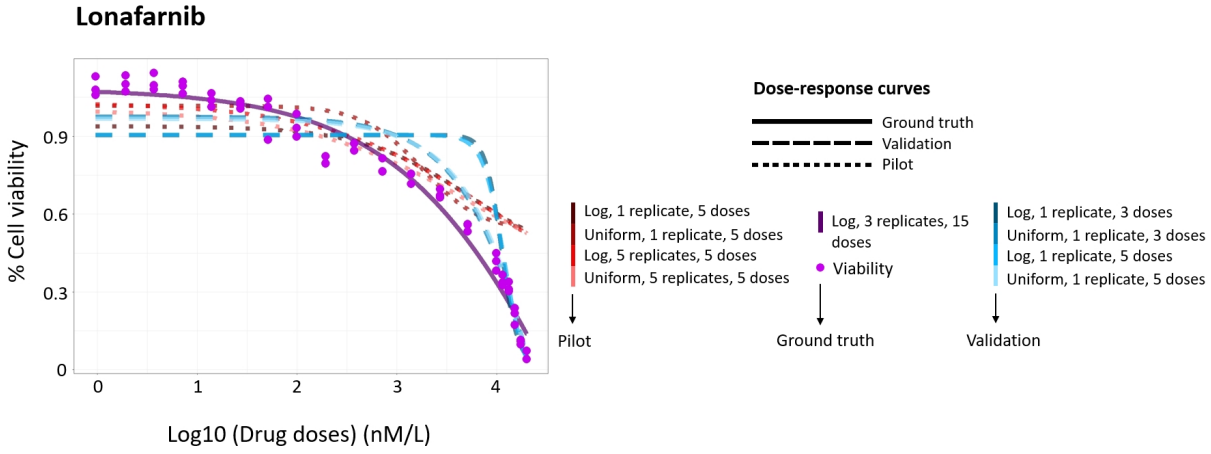

Figure 5.16: Validation assay analysis of design performance for Lonafarnib.

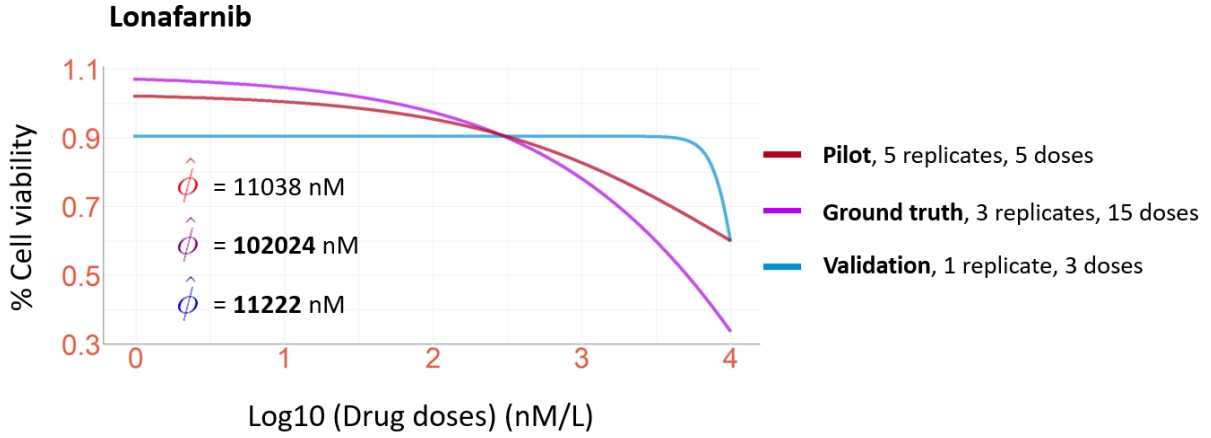

Figure 5.17: Comparison of dose-response curves determined by pilot, validation, and the ground truth on the log dose interval for Lonafarnib.

5.2.6 Carfilzomib and Bortezomib

Carfilzomib is a proteasome inhibitor permanently binding and changing the properties of the proteasome, thereby regulating cell-cycle progression and apoptosis (Stewart et al. 2015). Bortezomib is another protease inhibitor with the theoretical expectation of impeding tumor MM cell lines growth. Bortezomib has various mechanisms of action,

for instance being a key regulator of cyclin kinases or inhibiting nuclear factors (David et al. 2005). Both drugs are highly potent in MM cell lines such that with just a dosage of 1 nM or below it, is able to effectively inhibit more than half of MM cells no matter the interval at which initial doses were chosen (Figure 5.18). However, both the pilot and the two-step design were not able to infer accurate DR curves either in log or uniform intervals. For both the Carfilzomib and Bortezomib drugs, the estimated $\hat{\phi}$ for the log interval were consistently superior than for uniform doses (high deviation of $\hat{\phi}$ from the ground truth (purple curve)) (Figure 5.18, **A** and **B**). Indeed, the dose selection algorithm proposes drug doses below the threshold of 1 nM for such cases since the ϕ value is below the first drug dosage from pilot. Intuitively, when doses are in a log scale, DR inferences are superior than for doses in uniform format due to the fact that the ϕ value is much below the first dose from pilot, so doses above it are non-relevant which affects negatively DR inferences. Importantly, the two-step design with the addition of 3 doses was able to correct DR inferences in the curve fitting step for doses in uniform scale ($\hat{\phi}$ of 14.7 nM to 8 nM and 15.8 nM to 9.2 nM for Carfilzomib and Bortezomib, respectively). These results suggest two conclusions: (**1**) the two-step design is able to correct high inference biases from the pilot experiment for uniform doses and (**2**) reformulation of the proposed design is required for such dose-response scenarios where highly effective compounds inhibit cells with the first initial concentration. Finally, relative errors of AUC for the two-step experimental design were lower for both doses in different intervals (0.27 to 0.89 from pilot in log scale, and 58 to 173 from pilot in uniform scale, respectively for Carfilzomib). Strikingly, the relative errors of ϕ from the two-step design were lower for tested doses in uniform interval but not for log. Similar results of both metrics for Bortezomib were found where it was easier to outperform pilot in the uniform interval but similar relative errors in log scale.

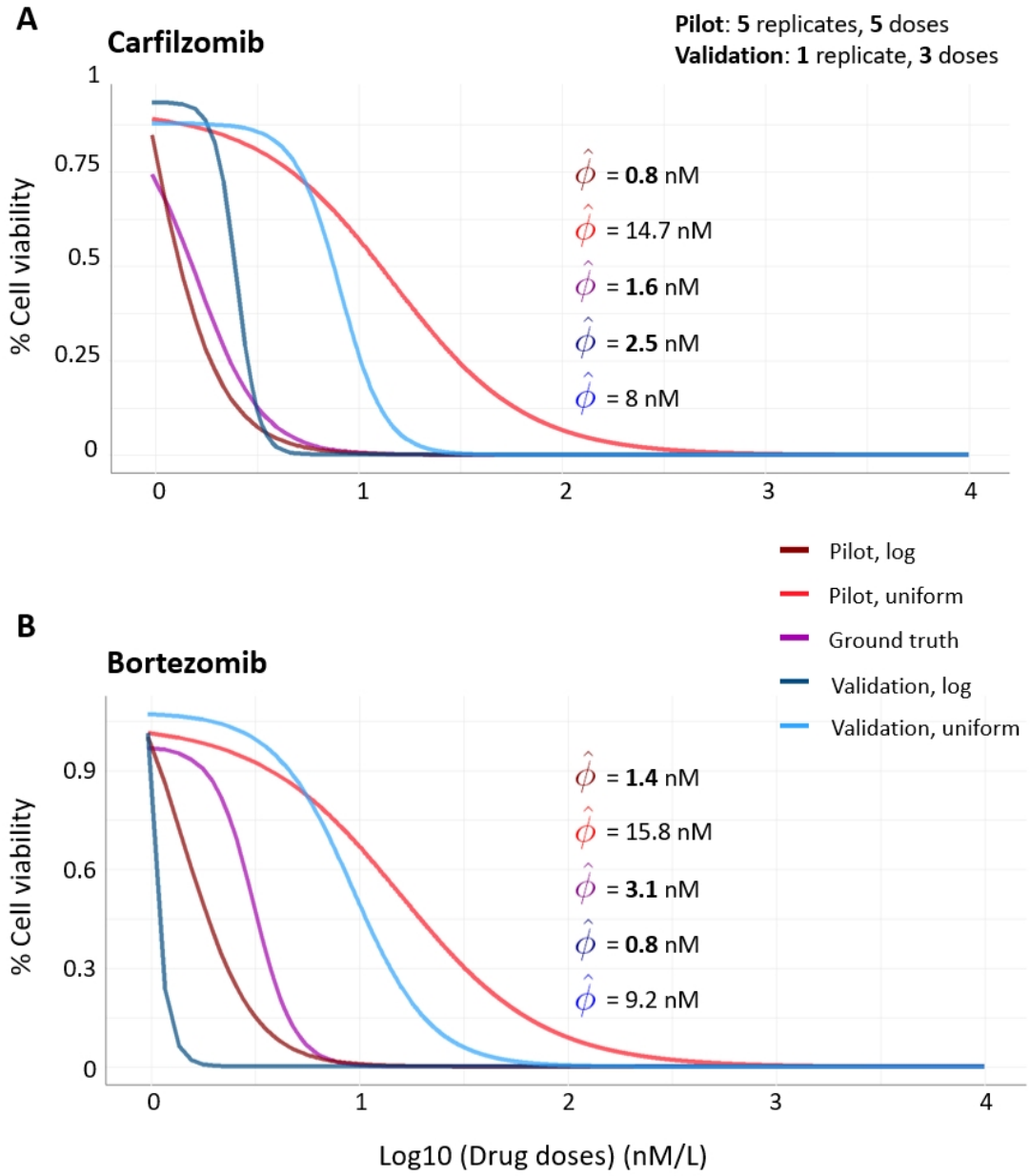

Figure 5.18: Two-step design cell viability analysis of Carfilzomib and Bortezomib.

Chapter 6

Discussion

In this thesis, a novel drug dose selection algorithm and experimental design was proposed to reduce biased DR curve inference evaluated in both simulation and experiment studies. Indeed, reproducibility (different lab technicians perform the same assay in distinct conditions) and replicability (the same experiment conducted several times - i.e., replicates), being essential cell-based assays, need to be controlled and standardized in the experimental design stage. Only through global efforts that the lack of scientific transparency and inconsistent reproducibility can be achieved.

A large-scale simulation workflow was adopted to first validate the proposed two-step design prior to the drug sensitivity assays. The outcomes of this simulation study were useful to acknowledge first, the critical regions of the DR curve and to propose a new dose selection algorithm (Figures 4.3, and 4.5). Across different values of ϕ, the addition of replicates was unable to reduce bias in the curve fitting step (Figure 5.1). This is surprising as replicates are standard protocols in a typical cell-based assay. In fact, no matter how many replicates are added, if the drug doses do not cover the critical regions of the DR curve (i.e., IC50), the inference will not be improved significantly. Based on such findings, the design was evaluated in a massive simulation workflow with many parameters (Table 4.1) governed by critical parameters of the 4-parametric logistic

model (Equation 2.1) based on their underlying behaviours (Figure 2.4). The results suggested that, by adopting a two-step experiment procedure, not only the curve bias is reduced but also the relative errors of the common drug sensitivity metrics ϕ and AUC (Figures 5.2, 5.3, and 5.4, respectively). Furthermore, based on the strategy, the distribution for absolute mean of bias across all parameters of the study depicted how highly effective the proposed design is in inferring DR relationship (Figure 5.7). Lastly, a full-scale statistical assessment for absolute mean of bias, median RE of ϕ, and mean RE of AUC confirmed the previously reported results, the two-step design outperformed the pilot experiment across different metrics (Table 5.1).

Validating the strategy required first data from the first round of experiments (denoted as pilot). In total, initial drug sensitivity assays were conducted in 6 multi-well culture plates and 2 for validation for MM cell lines and 7 chemical compounds that governed distinct ϕ and DR behaviour scenarios. For instance, it was expected that both Carfilzomib and Bortezomib drugs inhibit MM cells with the first dose at 1nM concentration. However, it did not improve DR estimation for doses in both log and uniform intervals (Figure 5.18). The expectation of the effect of Vorinostat, AZD4547 and Axitinib was such that by the third or fourth dose in log scale, cells would be inhibited by half (Figures 5.8, 5.10, and 5.12). For Ruxolitinib and Lonafarnib, the $\hat{\phi}$ estimations were the most challenging to correct, due to the fact that the IC50 is ill-defined with the pilot experiment tested doses, and thus, the algorithm is not able to improve the DR estimation that easily (Figures 5.14, and 5.16). In this study, to evaluate how accurately the two-step experimental design was able to estimate the *true* DR curve compared to the traditional one-step dosage design which involves multiple replicates, altogether, biological replicates (i.e., samples derived from distinct patients) for each drug over 15 concentration doses were added to the pilot assay plates and considered as the *ground truth*. Based on the pilot and validation assay results,

important outcomes highlighted both the efficiency and limitations of the proposed approach. First, it was shown that the proposed two-step design outperformed pilot experiment with high amounts of replicates for Vorinostat (Figure 5.9), AZD4547 (Figure 5.11), Axitinib (Figure 5.13), Ruxolitinib (Figure 5.15), and Lonafarnib (Figure 5.17). Doses proposed by the algorithm were able to consistently cover the critical regions of the *true* ϕ value thereby improving the DR inferences (although in some cases was more challenging such as for Ruxolitinib and Lonafarnib). Notwithstanding, from the analysis of the *ground truth* (i.e., 15 doses and 5 replicates), pilot, and validation: Lonafarnib proved to be both the most challenging drug to fit a DR curve and the most less potent compound with a IC50 surrounding 102024 nM. In general, the relative errors of AUC and ϕ suggest that the two-step design outperforms a pilot experiment with 5 replicates which is already substantially high. In summary, relative error of AUC for the design was better for Vorinostat, Axitinib, Ruxolitinib, Carfilzomib, and Bortezomib for doses within the uniform interval; and the relative error of ϕ for Vorinostat, AZD4547, Ruxolitinib, Lonafarnib, Bortezomib for doses within the uniform interval. Albeit with some challenges and limitations of the design, the implicit performance over a single experiment with high amounts of replicates is evident.

Indeed, the two-step experimental design and dose selection algorithm were developed under the assumptions from the 4-parameter logistic model (VØlund 1978). Hence, in the event of unmet expectations (i.e., a dose-response not representing a sigmoidal behaviour), poses limitations in the curve fitting step of drug sensitivity data analysis, affecting the IC50 and AUC metrics in the determination of the drug efficacy. In such cases, the use of non-parametric models could be an option (Amiryousefi et al. 2021). Another challenge of the design could be in a setting where patient samples need to be screened against chemical compounds, and it is not advisable nor possible in most cases to refreeze patient samples and conduct two rounds of experiments. Future steps include

optimizing the experimental design to tackle unsuccessful validation outcomes such as for the Carfilzomib and Bortezomib drugs (Figure 5.18), and to provide the new design as an R package or a web application for drug screening researchers.

Lastly, it is important to accurately identify which drug sensitivity metric to adopt when analysing dose-response data depending on the experimental design. This thesis focused on absolute mean of bias, therefore for the proposed two-step experimental design, bias should be the main metric of reproducibility. However, it was shown that the relative errors of both AUC and ϕ were generally improved, which indicates that the use of IC50 and AUC metrics for drug sensitivity screening can be also an option. In the future, it is important to consider all of the three metrics (i.e., the bias of viability, and errors of both IC50 and AUC) to have a complete overview of the data reproducibility for drug screening assays.

Chapter 7

Conclusion

Despite significant efforts highlighting the importance of data reliability apace with reaching consensus standards in cell-based assays (Larsson et al. 2020; Niepel et al. 2019), applicable strategies to reduce erroneous inferences in the dose-response curve fitting stage have been lacking. An experimental design was proposed to improve the reproducibility of drug sensitivity assays through efficient drug dose selection based on the initial experimental data. Thus, the robustness of current drug sensitivity and resistance metrics is expected to increase. In the event of lack of drug response data for the validation step, the design will be able to suggest proper dose concentration ranges and amount of technical or biological replicates to employ depending on the available resources. Taken together, the results of this thesis study suggests an improvement on the reproducibility in drug sensitivity assays and warrant further research to, ideally, minimize poor experimental designs and improve clinical translation from drug sensitivity screening.

Bibliography

Adam, Ishag et al. (2018). "Efficacy and safety of artemisinin-based combination therapy for uncomplicated Plasmodium falciparum malaria in Sudan: a systematic review and meta-analysis". In: *Malaria journal* 17.1, pp. 1–8.

Ahmed, S Ansar et al. (1994). "A new rapid and simple non-radioactive assay to monitor and determine the proliferation of lymphocytes: an alternative to [3H] thymidine incorporation assay". In: *Journal of immunological methods* 170.2, pp. 211–224.

Akaike, Hirotugu (1974). "A new look at the statistical model identification". In: *IEEE transactions on automatic control* 19.6, pp. 716–723.

Alexandrov, Ludmil B et al. (2020). "The repertoire of mutational signatures in human cancer". In: *Nature* 578.7793, pp. 94–101.

Amiryousefi, Ali et al. (2021). "The ENDS of assumptions; an online tool for the Epistemic Nonparametric Drug-response Scoring". In: *bioRxiv*.

Arrowsmith, John (2011). "Trial watch: Phase II failures: 2008-2010." In: *Nature reviews. Drug discovery* 10.5, pp. 328–329.

Baker, Monya (2016). "Biotech giant publishes failures to confirm high-profile science". In: *Nature News* 530.7589, p. 141.

Barretina, Jordi et al. (2012). "The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity". In: *Nature* 483.7391, pp. 603–607.

Basu, Amrita et al. (2013). "An interactive resource to identify cancer genetic and lineage dependencies targeted by small molecules". In: *Cell* 154.5, pp. 1151–1161.

Baty, Florent et al. (2015). "A toolbox for nonlinear regression in R: the package nlstools". In: *Journal of Statistical Software* 66, pp. 1–21.

Berenbaum, Morris C (1989). "What is synergy?" In: *Pharmacological reviews* 41.2, pp. 93–141.

Bliss, Chester I (1939). "The toxicity of poisons applied jointly 1". In: *Annals of applied biology* 26.3, pp. 585–615.

Bolomsky, Arnold et al. (2021). "Heterogeneous modulation of Bcl-2 family members and drug efflux mediate MCL-1 inhibitor resistance in multiple myeloma". In: *Blood advances* 5.20, pp. 4125–4139.

Bornkamp, Bjoern et al. (2010). "DoseFinding: Planning and analyzing dose finding experiments". In: *R package version 0.4-1*.

Bornkamp, Björn et al. (2009). "MCPMod: An R package for the design and analysis of dose-finding studies". In: *Journal of Statistical Software* 29, pp. 1–23.

Bornkamp, Björn et al. (2011). "Response-adaptive dose-finding under model uncertainty". In: *The Annals of Applied Statistics*, pp. 1611–1631.

Bunker, Alex et al. (2020). "Mechanistic understanding from molecular dynamics simulation in pharmaceutical research 1: drug delivery". In: *Frontiers in Molecular Biosciences*, p. 371.

Bushway, Paul J et al. (2010). "Hybrid median filter background estimator for correcting distortions in microtiter plate data". In: *ASSAY and Drug Development Technologies* 8.2, pp. 238–250.

Chan, Grace Ka Yan et al. (2013). "A simple high-content cell cycle assay reveals frequent discrepancies between cell number and ATP and MTS proliferation assays". In: *PloS one* 8.5, e63583.

Collignon, A et al. (2020). "A chemogenomic approach to identify personalized therapy for patients with relapse or refractory acute myeloid leukemia: results of a prospective feasibility study". In: *Blood cancer journal* 10.6, pp. 1–11.

Commo, Frederic et al. (2016). "R package nplr n-parameter logistic regressions". In: *V. 0.1–7*.

Consortium, Cancer Cell Line Encyclopedia et al. (2015). "Pharmacogenomic agreement between two cancer cell line data sets". In: *Nature* 528.7580, pp. 84–87.

David, Ebenezer et al. (2005). "The combination of the farnesyl transferase inhibitor lonafarnib and the proteasome inhibitor bortezomib induces synergistic apoptosis in human myeloma cells that is associated with down-regulation of p-AKT". In: *Blood* 106.13, pp. 4322–4329.

Dong, Zuoli et al. (2015). "Anticancer drug sensitivity prediction in cell lines from baseline gene expression through recursive feature selection". In: *BMC cancer* 15.1, pp. 1–12.

Downward, Julian (2003). "Targeting RAS signalling pathways in cancer therapy". In: *Nature reviews cancer* 3.1, pp. 11–22.

Ekstrøm, Claus Thorn (2020). *Package 'MESS'*.

Errington, Timothy M et al. (2014). "Science forum: An open investigation of the reproducibility of cancer biology research". In: *Elife* 3, e04333.

Fallahi-Sichani, Mohammad et al. (2013). "Metrics other than potency reveal systematic variation in responses to cancer drugs". In: *Nature chemical biology* 9.11, pp. 708–714.

Findlay, John WA et al. (2007). "Appropriate calibration curve fitting in ligand binding assays". In: *The AAPS journal* 9.2, E260–E267.

Fletcher, Roger (2013). *Practical methods of optimization*. John Wiley & Sons.

Forbes, Simon A et al. (2017). "COSMIC: somatic cancer genetics at high-resolution". In: *Nucleic acids research* 45.D1, pp. D777–D783.

Gavine, Paul R et al. (2012). "AZD4547: an orally bioavailable, potent, and selective inhibitor of the fibroblast growth factor receptor tyrosine kinase family". In: *Cancer research* 72.8, pp. 2045–2056.

Gerhard, Daniel et al. (2017). "Marginalization in nonlinear mixed-effects models with an application to dose-response analysis". In: *arXiv preprint arXiv:1707.02502*.

Graul, Ann I et al. (2009). "Overcoming the obstacles in the pharma/biotech industry: 2008 update". In: *Drug News Perspect* 22.1, p. 39.

Graul, Ann I et al. (2010). "Overcoming the obstacles in the pharma/biotech industry: 2009 update". In: *Drug News Perspect* 23.1, pp. 48–63.

Gupta, Abhishekh et al. (2020). "A normalized drug response metric improves accuracy and consistency of anticancer drug sensitivity quantification in cell-based screening". In: *Communications biology* 3.1, pp. 1–12.

Hafner, Marc et al. (2016). “Growth rate inhibition metrics correct for confounders in measuring sensitivity to cancer drugs”. In: *Nature methods* 13.6, pp. 521–527.

Haibe-Kains, Benjamin et al. (2013). “Inconsistency in large pharmacogenomic studies”. In: *Nature* 504.7480, pp. 389–393.

Harris, Leonard A et al. (2016). “An unbiased metric of antiproliferative drug effect in vitro”. In: *Nature methods* 13.6, pp. 497–500.

Haverty, Peter M et al. (2016). “Reproducible pharmacogenomic profiling of cancer cell line panels”. In: *Nature* 533.7603, pp. 333–337.

Heiser, Laura M et al. (2012). “Subtype and pathway specific responses to anticancer compounds in breast cancer”. In: *Proceedings of the National Academy of Sciences* 109.8, pp. 2724–2729.

Holbeck, Susan L et al. (2017). “The National Cancer Institute ALMANAC: a comprehensive screening resource for the detection of anticancer drug pairs with enhanced therapeutic activity”. In: *Cancer research* 77.13, pp. 3564–3576.

Ianevski, Aleksandr et al. (2017). “SynergyFinder: a web application for analyzing drug combination dose–response matrix data”. In: *Bioinformatics* 33.15, pp. 2413–2415.

Jaaks, Patricia et al. (2022). “Effective drug combinations in breast, colon and pancreatic cancer cells”. In: *Nature* 603.7899, pp. 166–173.

Jaiswal, Alok et al. (2019). “Integrated analysis of drug sensitivity and selectivity to predict synergistic drug combinations and target coaddictions in cancer”. In: *Systems Chemical Biology*. Springer, pp. 205–217.

Jakštys, Baltramiejus et al. (2015). “Different cell viability assays reveal inconsistent results after bleomycin electrotransfer in vitro”. In: *The Journal of membrane biology* 248.5, pp. 857–863.

Jones, Laurie J et al. (2001). “Sensitive determination of cell number using the CyQUANT® cell proliferation assay”. In: *Journal of immunological methods* 254.1-2, pp. 85–98.

Julkunen, Heli et al. (2020). “Leveraging multi-way interactions for systematic prediction of pre-clinical drug combination effects”. In: *Nature communications* 11.1, pp. 1–11.

Kahm, Matthias et al. (2010). “Grofit: fitting biological growth curves”. In: *Nature Precedings*, pp. 1–1.

Kim, Sunghwan et al. (2021). “PubChem in 2021: new data content and improved web interfaces”. In: *Nucleic acids research* 49.D1, pp. D1388–D1395.

Kitano, Hiroaki (2002). “Computational systems biology”. In: *Nature* 420.6912, pp. 206–210.

Kuo, Wen-Lin et al. (2009). “A systems analysis of the chemosensitivity of breast cancer cells to the polyamine analogue PG-11047”. In: *BMC medicine* 7.1, pp. 1–11.

Larsson, Peter et al. (2020). “Optimization of cell viability assays to improve replicability and reproducibility of cancer drug sensitivity screens”. In: *Scientific reports* 10.1, pp. 1–12.

Lee, Daniel et al. (2000). “Algorithms for non-negative matrix factorization”. In: *Advances in neural information processing systems* 13.

Li, Jun et al. (2017). “Characterization of human cancer cell lines by reverse-phase protein arrays”. In: *Cancer cell* 31.2, pp. 225–239.

Lin, Lihui et al. (2020). “Ex-vivo drug testing predicts chemosensitivity in acute myeloid leukemia”. In: *Journal of Leukocyte Biology* 107.5, pp. 859–870.

Liu, Dong C et al. (1989). “On the limited memory BFGS method for large scale optimization”. In: *Mathematical programming* 45.1, pp. 503–528.

Loewe, S (1953). “The problem of synergism and antagonism of combined drugs”. In: *Arzneimittelforschung* 3, pp. 285–290.

Macdougall, James (2006). “Analysis of dose–response studies—E max model”. In: *Dose finding in drug development*. Springer, pp. 127–145.

Majumder, Muntasir Mamun et al. (2017). “Identification of precision treatment strategies for relapsed/refractory multiple myeloma by functional drug sensitivity testing”. In: *Oncotarget* 8.34, p. 56338.

Malani, D et al. (2017). “Enhanced sensitivity to glucocorticoids in cytarabine-resistant AML”. In: *Leukemia* 31.5, pp. 1187–1195.

Malani, Disha et al. (2021). “Implementing a functional precision medicine tumor board for acute myeloid leukemia”. In: *Cancer Discovery*.

Malyutina, Alina et al. (2019). “Drug combination sensitivity scoring facilitates the discovery of synergistic and efficacious drug combinations in cancer”. In: *PLoS computational biology* 15.5, e1006752.

Malyutina, Alina et al. (2021). “drda: An R package for dose-response data analysis”. In: *bioRxiv*.

Manuvakhova, Marina S et al. (2011). “Identification of novel small molecule activators of nuclear factor-κb with neuroprotective action via high-throughput screening”. In: *Journal of neuroscience research* 89.1, pp. 58–72.

Markossian, Sarine et al. (2021). *Assay Guidance Manual for Drug Discovery: Robust or Go Bust*.

Mascarenhas, John et al. (2012). “Ruxolitinib: the first FDA approved therapy for the treatment of myelofibrosis”. In: *Clinical cancer research* 18.11, pp. 3008–3014.

McMillian, MK et al. (2002). “An improved resazurin-based cytotoxicity assay for hepatic cells”. In: *Cell biology and toxicology* 18.3, pp. 157–173.

McTigue, Michele et al. (2012). “Molecular conformations, interactions, and properties associated with drug efficiency and clinical performance among VEGFR TK inhibitors”. In: *Proceedings of the National Academy of Sciences* 109.45, pp. 18281–18289.

Mendez, David et al. (2019). “ChEMBL: towards direct deposition of bioassay data”. In: *Nucleic acids research* 47.D1, pp. D930–D940.

Mosmann, Tim (1983). “Rapid colorimetric assay for cellular growth and survival: application to proliferation and cytotoxicity assays”. In: *Journal of immunological methods* 65.1-2, pp. 55–63.

Mpindi, John-Patrick et al. (2015). “Impact of normalization methods on high-throughput screening data with high hit rates and drug testing with dose–response data”. In: *Bioinformatics* 31.23, pp. 3815–3821.

Niepel, Mario et al. (2017). “Measuring cancer drug sensitivity and resistance in cultured cells”. In: *Current protocols in chemical biology* 9.2, pp. 55–74.

Niepel, Mario et al. (2019). “A multi-center study on the reproducibility of drug-response assays in mammalian cell lines”. In: *Cell systems* 9.1, pp. 35–48.

Niles, Andrew L et al. (2007). "A homogeneous assay to measure live and dead cells in the same sample by detecting different protease markers". In: *Analytical biochemistry* 366.2, pp. 197–206.

Nocedal, Jorge et al. (2006). "Springer series in operations research and financial engineering". In: *Numerical optimization*. Springer New York.

O'brien, John et al. (2000). "Investigation of the Alamar Blue (resazurin) fluorescent dye for the assessment of mammalian cell cytotoxicity". In: *European journal of biochemistry* 267.17, pp. 5421–5426.

O'Neil, Jennifer et al. (2016). "An unbiased oncology compound screen to identify novel combination strategies". In: *Molecular cancer therapeutics* 15.6, pp. 1155–1162.

Ouyang, Defang et al. (2015). "Introduction to computational pharmaceutics". In: *Computational Pharmaceutics*, pp. 1–5.

Pemovska, Tea et al. (2013). "Individualized systems medicine strategy to tailor treatments for patients with chemorefractory acute myeloid leukemia". In: *Cancer discovery* 3.12, pp. 1416–1429.

Pemovska, Tea et al. (2015). "Axitinib effectively inhibits BCR-ABL1 (T315I) with a distinct binding conformation". In: *Nature* 519.7541, pp. 102–105.

Pierre, Robert V (2002). "Peripheral blood film review: the demise of the eyecount leukocyte differential". In: *Clinics in laboratory medicine* 22.1, pp. 279–297.

Pietarinen, Paavo O et al. (2017). "Differentiation status of primary chronic myeloid leukemia cells affects sensitivity to BCR-ABL1 inhibitors". In: *Oncotarget* 8.14, p. 22606.

Pietarinen, PO et al. (2015). "Novel drug candidates for blast phase chronic myeloid leukemia from high-throughput drug sensitivity and resistance testing". In: *Blood cancer journal* 5.5, e309–e309.

Pinheiro, Jose et al. (2007). "Linear and nonlinear mixed effects models". In: *R package version* 3.57, pp. 1–89.

Pinheiro, José et al. (2014). "Model-based dose finding under model uncertainty using general parametric models". In: *Statistics in medicine* 33.10, pp. 1646–1661.

Posner, Bruce A (2005). "High-throughput screening-driven lead discovery: meeting the challenges of finding new therapeutics." In: *Current opinion in drug discovery & development* 8.4, pp. 487–494.

Potdar, Swapnil et al. (2020). "Breeze: an integrated quality control and data analysis application for high-throughput drug screening". In: *Bioinformatics* 36.11, pp. 3602–3604.

Quent, Verena MC et al. (2010). "Discrepancies between metabolic activity and DNA content as tool to assess cell proliferation in cancer research". In: *Journal of cellular and molecular medicine* 14.4, pp. 1003–1013.

Ramirez, Michael et al. (2016). "Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells". In: *Nature communications* 7.1, pp. 1–8.

Richards, FJ (1959). "A flexible growth function for empirical use". In: *Journal of experimental Botany* 10.2, pp. 290–301.

Ritz, Christian et al. (2015). "Dose-response analysis using R". In: *PloS one* 10.12, e0146021.

Root, David E et al. (2003). "Detecting spatial patterns in biological array experiments". In: *Journal of biomolecular screening* 8.4, pp. 393–398.

Safikhani, Zhaleh et al. (2016a). "Assessment of pharmacogenomic agreement". In: *F1000Research* 5.

Safikhani, Zhaleh et al. (2016b). "Revisiting inconsistency in large pharmacogenomic studies". In: *F1000Research* 5.

Schwarz, Gideon (1978). "Estimating the dimension of a model". In: *The annals of statistics*, pp. 461–464.

Seashore-Ludlow, Brinton et al. (2015). "Harnessing connectivity in a large-scale small-molecule sensitivity dataset". In: *Cancer discovery* 5.11, pp. 1210–1223.

Shoemaker, Robert H (2006). "The NCI60 human tumour cell line anticancer drug screen". In: *Nature Reviews Cancer* 6.10, pp. 813–823.

Siegel, DS et al. (2014). "Vorinostat in combination with lenalidomide and dexamethasone in patients with relapsed or refractory multiple myeloma". In: *Blood cancer journal* 4.2, e182–e182.

Silva, Gabriela et al. (2013). “Vorinostat induces apoptosis and differentiation in myeloid malignancies: genetic and molecular mechanisms”. In: *PloS one* 8.1, e53766.

Skaga, Erlend et al. (2019). “Feasibility study of using high-throughput drug sensitivity testing to target recurrent glioblastoma stem cells for individualized treatment”. In: *Clinical and translational medicine* 8.1, pp. 1–15.

Slater, TF et al. (1963). “Studies on succinate-tetrazolium reductase systems: III. Points of coupling of four different tetrazolium salts III. Points of coupling of four different tetrazolium salts”. In: *Biochimica et biophysica acta* 77, pp. 383–393.

Smirnov, Petr et al. (2018). “PharmacoDB: an integrative database for mining in vitro anticancer drug screening studies”. In: *Nucleic acids research* 46.D1, pp. D994–D1002.

Steihaug, Trond (1983). “The conjugate gradient method and trust regions in large scale optimization”. In: *SIAM Journal on Numerical Analysis* 20.3, pp. 626–637.

Stewart, A Keith et al. (2015). “Carfilzomib, lenalidomide, and dexamethasone for relapsed multiple myeloma”. In: *New England Journal of Medicine* 372.2, pp. 142–152.

Sun, Wei et al. (2017). “Synergistic drug combination effectively blocks Ebola virus infection”. In: *Antiviral research* 137, pp. 165–172.

Szwajda, Agnieszka et al. (2015). “Systematic mapping of kinase addiction combinations in breast cancer cells by integrating drug sensitivity and selectivity profiles”. In: *Chemistry & Biology* 22.8, pp. 1144–1155.

Tang, Jing et al. (2013). “Target inhibition networks: predicting selective combinations of druggable targets to block cancer survival pathways”. In: *PLoS computational biology* 9.9, e1003226.

Tang, Jing et al. (2018). “Drug target commons: a community effort to build a consensus knowledge base for drug-target interactions”. In: *Cell chemical biology* 25.2, pp. 224–229.

Tanoli, Ziaurrehman et al. (2021). “Exploration of databases and methods supporting drug repurposing: a comprehensive survey”. In: *Briefings in bioinformatics* 22.2, pp. 1656–1678.

Tate, John G et al. (2019). “COSMIC: the catalogue of somatic mutations in cancer”. In: *Nucleic acids research* 47.D1, pp. D941–D947.

Tolliday, Nicola (2010). "High-throughput assessment of mammalian cell viability by determination of adenosine triphosphate levels". In: *Current protocols in chemical biology* 2.3, pp. 153–161.

Tyner, Jeffrey W et al. (2013). "Kinase pathway dependence in primary human leukemias determined by rapid inhibitor screening". In: *Cancer research* 73.1, pp. 285–296.

Vis, Daniel J et al. (2016). "Multilevel models improve precision and speed of IC50 estimates". In: *Pharmacogenomics* 17.7, pp. 691–700.

Visser, Ubbo et al. (2011). "BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results". In: *BMC bioinformatics* 12.1, pp. 1–16.

VØlund, Aage (1978). "Application of the four-parameter logistic model to bioassay: comparison with slope ratio and parallel line models". In: *Biometrics*, pp. 357–365.

Weinstein, John N (2012). "Cell lines battle cancer". In: *Nature* 483.7391, pp. 544–545.

Wilkinson, Mark D et al. (2016). "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1, pp. 1–9.

Yadav, Bhagwan et al. (2014). "Quantitative scoring of differential drug sensitivity for individually optimized anticancer therapies". In: *Scientific reports* 4.1, pp. 1–10.

Yadav, Bhagwan et al. (2015). "Searching for drug synergy in complex dose–response landscapes using an interaction potency model". In: *Computational and structural biotechnology journal* 13, pp. 504–513.

Yang, Wanjuan et al. (2012). "Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells". In: *Nucleic acids research* 41.D1, pp. D955–D961.

Yoshida, Mitsuzi et al. (1975). "Action of 5-fluorocyclocytidine on cultured L-5178Y cells". In: *GANN Japanese Journal of Cancer Research* 66.5, pp. 561–564.

Zagidullin, Bulat et al. (2019). "DrugComb: an integrative cancer drug combination data portal". In: *Nucleic acids research* 47.W1, W43–W51.

Zander Balderud, Linda et al. (2015). "Using the BioAssay Ontology for analyzing high-throughput screening data". In: *Journal of biomolecular screening* 20.3, pp. 402–415.

Zhang, Ji-Hu et al. (1999). "A simple statistical parameter for use in evaluation and validation of high throughput screening assays". In: *Journal of biomolecular screening* 4.2, pp. 67–73.

Zhang, Xiaohua Douglas (2007). "A pair of new statistical parameters for quality control in RNA interference high-throughput screening assays". In: *Genomics* 89.4, pp. 552–561.

Zheng, Shuyu et al. (2021). "DrugComb update: a more comprehensive drug sensitivity data repository and analysis portal". In: *Nucleic acids research* 49.W1, W174–W184.

Zheng, Shuyu et al. (2022). "SynergyFinder Plus: Toward Better Interpretation and Annotation of Drug Combination Screening Datasets". In: *Genomics, Proteomics & Bioinformatics*.

Zheng, Wei et al. (2013). "Phenotypic screens as a renewed approach for drug discovery". In: *Drug discovery today* 18.21-22, pp. 1067–1073.

Zur Wiesch, Pia Abel et al. (2011). "Population biological principles of drug-resistance evolution in infectious diseases". In: *The Lancet infectious diseases* 11.3, pp. 236–247.