Drug Development for Patients

Personalised Dosing For All

Alan Maloney

1 May 2023

Table of contents

Welcome

Modern drug development is failing patients; it is painfully outdated and desperately needs to change.

This book explains why we must put **individual** patient outcomes at the centre of drug development and regulatory approval.

It explains why "one-size-fits-all" dosing is unscientific and incorrect; our (heterogeneous) patients often suffer terribly because we persist with the same "simple" fixed-dose regimen for all. We must refocus drug development to be truly "patient centric" and deliver **Personalised Dosing** (i.e. dose individualisation). By chapter 4, you will understand that our patients are not "fields".

The recent draft FDA guidance on dose optimisation in oncology and **Project Optimus** both aim to improve dosing in oncology; this is excellent, but the continued emphasis on "one-size-fits-all" dosing is incorrect. Thus I have decided to release this (work in progress) book will the goal to both convince and motivate all stakeholders (patients, regulators, pharma, payers) to support a patient focused, **Drug Development For Patients** paradigm.

These FDA initiatives must state that smart, science-based dose titration algorithms that deliver **Personalised Dosing** will lead to much better individual patient outcomes than any fixed-dose regimens. **If we truly care about obtaining the best individual patient outcomes, we must embrace Personalised Dosing across all therapeutic areas.**

After presenting the clear and compelling scientific arguments, I will ask you in the conclusion:

- Are you a part of the problem, continuing to accept fixed-dose regimens because you are very familiar with "simple", even though this leads to worse patient outcomes?

- Or are you part of the solution, prepared to embrace Personalised Dosing because you truly care and want the best possible outcomes for our patients?

This is your choice. I hope by the end of this book you will be convinced to choose **and** actively advocate for the latter; many already do, but we need your voice too!

Hopefully we can all pull together and make it happen.

Happy reading!

Al Maloney

1 Introduction

This book aims to explain why we need to put individual patient outcomes at the centre of drug development and regulatory approval. To achieve this, we need to fully recognise heterogeneity between patients, and the central role played by dose.

Drug development is about solving a puzzle. We have patients, and we have a drug.

Our goal is to work out how best to treat each and every patient using the drug.

Stated more mathematically, this is an optimisation problem, where we aim to maximise the outcomes for patients using the drug optimally. With this goal, all clinical trials in the development of a drug should be focussed on efficiently and quickly solving this puzzle.

Higher doses will, in most cases, lead to greater effects that can be both positive (efficacy) and negative (tolerability/safety) for the patient. As noted by Paracelsus in 1538:

> "*Only the dose makes the poison*"

We must therefore determine how best to balance these opposing effects. Too low a dose, and the patient will not benefit as much as they could; too high a dose, and the patient will experience unnecessarily severe tolerability/safety issues. Crucially, these dose-response relationships, and hence the optimal dose, will **differ** between our heterogeneous patients. Whilst the best dose for Emma may be 10 mg, for Casper it may be 100 mg.

Drug development is, and always will be, a scientific learning exercise. Regulatory approval should be based on understanding both **Population** and **Individual** dose-response (D-R) relationships across multiple efficacy and tolerability/safety endpoints, using the totality of the accrued data from all trials. With a science-based dose titration algorithm, we can achieve the best dose for each and every patient; we can truly deliver **Personalised Dosing**.

The current way we develop drugs is painfully crude and does not achieve our goal. Consequently, far too many patients are given the wrong dose for them.

This book will explain the main problems, and propose solutions. Broadly speaking, the solutions fall into 5 primary themes:

1) Understanding the goal

2) Understanding the science

3) Understanding dose-response models/modelling

4) Changes needed within the regulatory agencies

5) Changes needed within the pharmaceutical industry

Despite the good intentions of all involved (myself included), we must recognise that the current "status quo" is wholly unacceptable for modern drug development; we can, and should, do so much better. Rather than seek to blame anyone or anything, we must simply acknowledge what is truly important, and then quickly move forward in a more scientific manner, where optimising patient outcomes is central to everything we do.

The above themes will appear throughout the book. Although it is important to discuss some of the current problems and causes of why we are where we are, this book aims to be more inspirational, and to say where we should be. As an analogy, rather than "patch up" an old derelict house with antiquated infrastructure, we can design a brand new house, using modern science and technology. If we could redesign drug development from the "floor up" with modern, patient-centric, scientifically designed and analysed drug programs, what would they look like? Can we be brave, and swiftly move to an approach that is truly better for patients?

In parts, the text will be pointed and unapologetically direct. I prefer directness, even if it may be harsh; I hope it is never construed as arrogant, rude or condescending. I do not aim to offend anyone, but my anger and frustration at drug development programs that are not "patient-centric" is genuine. If we truly focus on individual patient outcomes, I honestly believe we can do so much better for patients.

I hope this introduction is appealing to people with an interest in drug development and improving patient outcomes, including regulatory agencies, the pharmaceutical industry, patients/patient advocacy groups and payers, and that you will continue reading to find out more .

Note: you can use the "back" button in your browser to return from a hyperlink, and can find all abbreviations defined in the Glossary.

2 My Motivation For This Book

Over the last 28 years I have worked in the pharmaceutical industry, either as an employee or as an external consultant. With a background in biostatistics and clinical pharmacology, my technical work/research has included:

- Design and analysis of clinical trials (D-R trials, head-to-head trials, proof of concept/principle trials, first in man trials etc.)

- Integrated dose-exposure-response (D-E-R) modelling and simulation across all trials and doses for sponsor decision making and regulatory submissions

- Integrated analyses across multiple trials, drugs and doses (so called Model-Based Meta-Analyses (MBMA)) for accurate/precise comparative effectiveness evaluations and probability of success simulations

- Optimal/adaptive trial designs for D-E-R modelling

- Combination therapy trial design/analysis

- Bayesian modelling

As a technical expert I use (all available) clinical trial data, mathematical models and simulations to provide integrated, quantitative analyses to guide drug development strategies. I have worked in most therapeutic areas, across phases 1-4 and all data types (continuous, categorical, count, survival etc.).

Fortunately for me, I have always found my work both interesting and challenging. In particular, I have strived to become an expert on D-R modelling, both by studying hundreds of D-R relationships across numerous types of clinical endpoints, and the use of appropriate mathematical models to describe such data.

However more importantly, I think it is essential to view our clinical trial individual patient data from the perspective of each individual patient. For example, that the dose they were assigned did not provide them with any meaningful benefit, or indeed caused them to experience real harms. Given my understanding of individual D-R relationships, I **know** we can do better, but we have to be open-minded when we think about what dose a patient should start with, and how we should best change the dose if needed.

I have also tried to understand the "big picture" of drug development, by considering the goals of others working in industry, regulatory authorities and reimbursement/payers. What is important to them, and what are their constraints? To select just a few observations:

- Drug manufacturing teams being asked to initiate production/testing of dose levels **before** any clinical data has been collected/analysed.

- Commercial teams wanting a "one-size-fits-all" dose, but being unaware of how this will negatively impact patient outcomes, adherence, and ultimately sales.

- Project teams being asked to run ever smaller and faster development programs, but still expected to make informed, scientifically sound, decisions.

- Internal regulatory teams wishing to follow the same path taken by other sponsors in earlier submissions (the safety strategy is the "same as last time" strategy).

- External regulatory teams being asked to approve/reject new drug applications when the sponsor only provides two trials with "P<0.05" for a primary endpoint for a single dose level, and "tables and listings" for all safety data.

- Payers being asked to reimburse expensive drugs for a patient, but having no idea whether the patient will actually respond.

The above gives a first insight into why companies do what they currently do; a "sprint" to get 1-2 fixed-dose regimens into phase 3, demonstrate superiority over placebo and seek approval based on whatever safety data was observed. Dose response, individual patient outcomes and dose individualisation are on the periphery/absent in this process; fundamentally such a strategy is devoid of the concept of **Personalised Dosing**.

To be clear then, changing drug development will be very hard, with multiple stakeholders and historical/legacy frameworks that unfortunately conspire to make a difficult job even more difficult. **However we must all endeavour to refocus drug development back to individual patient outcomes (are you with me?).**

As a young analyst, I had the good fortune to talk with the brilliant Lewis Sheiner. He was a visionary in how he viewed drug development through both the sharp lens of clinical pharmacology and sound statistical analysis. When we think about drug development, we might ask, "**What would Lewis Sheiner do?**" He was a passionate advocate of scientific debate and progressive thinking, and remains my idol and someone I aspire to be more like.

My experiences in drug development have therefore led me to want to write this book, in the hope that it will help improve how we discuss drug development, how we design and analyse the required clinical trials, and how regulators use this comprehensive evidence base to ensure drugs are used optimally for all patients.

> **!** Important
>
> I think to be a good drug developer you must truly empathise with each patient taking your drug. If a patient is not responding, or experiences safety/tolerability issues, we must "own" this failure, learn from it, and strive to do better going forward both for this patient and the next.

Such a continual improvement mindset will result in better outcomes for patients. Better outcomes for patients will also benefit payers and the pharmaceutical industry; all stakeholders will benefit from the proposed "roadmap" for drug development herein.

Although you may not agree with everything in this book, I hope it will promote honest dialogue and real momentum for change. If they are willing to instigate real change, I am optimistic that leaders of regulatory agencies, industry, patient advocacy groups and payers can work closely together to make it happen.

3 Why Patients Outcomes Must Come First

At the end of this chapter the reader will understand:

- The need to put the outcomes of each and every patient at the centre of everything we do in drug development.

- If we wish to improve individual patient outcomes, we must look beyond "average" responses from fixed-dose regimens.

- That **science-based dose titration algorithms** can enable **Personalised Dosing**; we seek to understand the best starting dose, how and when to titrate, and when to consider halting treatment.

It may seem odd to need to say this, but drug development needs to put the outcomes of each and every patient at the centre of everything we do. To repeat:

> **!** Important
>
> Drug development needs to put the outcomes of each and every patient at the centre of everything we do.

It is perhaps easy for those involved in drug development to become detached from the realities of how the dose regimen each patient receives changes their life, both for good and bad. Rather, we might see summary tables by treatment arm (e.g. placebo and drug) with the proportions of responders and non-responders (with a "significant" P value), and the proportions of patients with (serious) adverse events. At a superficial level the pharmaceutical company may argue "job done", and proceed to file for regulatory approval.

But if we pause for a moment, there are salient questions we should be asking. For example, consider the non-responders on drug; why did these patients not respond, or have an inadequate/poor response? Surely these patients should be investigated more thoroughly, since the dose regimen has failed them. Is seems wholly remiss to simply shrug our shoulders and conclude that the proposed dosing regimen "cures all but the incurable". The central role

of the dose regimen has been ignored. We can only conclude that **this** dose regimen failed these patients; this is very different from concluding there is no acceptable dosing regimen for these patients. Perhaps a patient dropped out due to unacceptable tolerability (consider a lower dose?) or had no tolerability issues, but just did not respond at this dose level (consider a higher dose?). Could we have not used **any** clinical endpoint or biomarker/imaging data to inform a **science-based dose titration strategy**, or simply asked the patient/physician team "Given your outcomes thus far, would you prefer to remain on the same dose, or consider a lower or higher dose?" Given our understanding of heterogeneity between patients in pharmacokinetics (PK), pharmacodynamics (PD), and their personal assessments of these effects, should we not plan to enable **Personalised Dosing** if we wish to put patient outcomes truly at the centre of drug development?

In oncology, a grade 2 toxicity of diarrhoea is defined as "Passing 4-6 more stools a day than your baseline" (generally only grade 3 and above are considered "severe" toxicities). How would we cope with such a debilitating experience? What impact would this have on our physiological and psychological states, or on our outlook towards future treatment cycles? Could we have worked that day, or fetched the kids from school? Behind a Kaplan-Meier overall survival curve in oncology, each downward step reflects a death of one or more patients. Just think about that. The dose didn't work and someone died, and people close to that patient will be devastated. Did we really do **everything** we could to get the drug to work for that patient? If a patient with cancer showed little or no positive response after initial treatment cycles (e.g. no/minimal reduction in tumour size), should we have not titrated the dose higher? Perhaps the drug concentrations in this patient were much lower than other patients who received the same dose, and hence a higher dose may have truly helped this patient.

At an FDA-ASCO Virtual Workshop meeting entitled "Getting the Dose Right: Optimizing Dose Selection Strategies in Oncology" in May 2022, one panellist described intra-patient dose titrations, like that described above, as "problematic" from an analysis perspective. Indeed, such data cannot always be analysed using simple methods. Put bluntly, I see it as much more "problematic" that patients are dying or experiences severe toxicities because they are not receiving the right dose for them. **We need to align trial designs and dosing regimens to put the outcomes of patients at the centre of everything we do, yet still generate meaningful data on how best to dose the drug**. More complex analyses are a small price to pay for better patient outcomes; we can, and must, do this.

I had the recent opportunity to listen to a remarkable lady, Jill Feldman. You can find more information about her both before and after her diagnosis and treatment for lung cancer (link 1 below) and see a nice example of her patient advocacy work (link 2).

https://www.iaslc.org/journey-unlike-any-other-jill-feldman

https://youtu.be/a1r4nTxbAE0

She described the horrendous list of adverse events she experienced when receiving her first drug regimen, even after two dose reductions. Jill also described how she "tried to look my

best" at physician visits to "make sure they didn't give up on me". She also described the dilemma she and her oncologist faced when she started a new drug regimen; go with the full dose, or "risk" a lower dose? The current drug label provided no guidance, since the right trials have not been done. Without this evidence, Jill went with the full dose, leading to what she described as "8 weeks of hell". Jill is not the first, and will not be the last, patient to be in this position. We need to provide a much better evidence base to help patients like Jill make informed decisions.

In summary, when we put patients outcomes first, the drug development paradigm becomes clear. We need to determine 3 things:

1) Given a patient's individual characteristics, what is the best **initial** drug and dosing regimen?

2) If/when the initial dosing regimen needs to be changed for efficacy and/or safety/tolerability, how best to do this; what is the **best science-based dose titration algorithm?** That is, based on clinical endpoints, biomarkers, imaging and/or patient reported outcomes (PROs), when should the dose be changed, and by how much.

3) Under what circumstances should the dosing regimen be halted? That is, there is no dose for the patient that has a sufficiently positive benefit-risk to justify continued dosing.

Although the above will not apply to every therapeutic area, the general principle should be clear; our goal is to obtain the best outcome for each and every patient via informed **Personalised Dosing**.

---

4 A Brief History Of Drug Development; The Good, The Bad And The Ugly

At the end of this chapter, the reader will understand:

- Why randomised controlled trials (RCTs) are considered the gold standard for assessing new drug regimens through their use of randomisation, blinding, and control arms.

- Understand how clinical drug development is based on a series of RCTs nominally categorised as phase 1, phase 2 and phase 3 trials.

- Why RCTs are not agricultural experiments, and patient are not fields! If we care about individual patient outcomes, we need to move beyond average outcomes based on the same fixed-dose regimen for all patients.

- Why drug development is an estimation problem, rather than a null hypothesis significance testing (NHST) problem.

- Why off-label drug use and some placebo controlled trials are unacceptable.

It is valuable to briefly reflect on the history of clinical trial design and drug development, and how this has evolved into the general drug development paradigm we see today; by looking back, we can see where we went right, and where we went wrong. Initially we will focus on the **good**.

Randomised controlled trials (RCT) are generally regarded as the gold standard by which we measure the benefit-risk of new experimental drug regimens. The first modern RCT is generally reported to be the 1948 General Medical Council trial that assessed the efficacy of an antibiotic, streptomycin, for the treatment of pulmonary tuberculosis [1]. The goal of this trial was to evaluate whether 500 mg of streptomycin given every 6 hours over 4-6 months improved survival versus standard of care.

Five key components that made this a "modern" trial was the application of:

- Inclusion/exclusion criteria

- Randomisation

- Blinding

- A control arm

- A formal statistical analysis

Their report listed how patients were required to satisfy a number of features/criteria that we would now refer to as inclusion/exclusion criteria. The authors added:

> *"Such closely defined features were considered indispensable for it was realized that no two patients have an identical form of the disease, and it was desired to eliminate as many of the obvious variations as possible."*

Although there is now a strong desire to ensure patients in modern clinical trials are truly representative of patients who may subsequently be given the drug, the aim of a clear and focussed experiment in a relatively homogeneous group of patients was reasonable in 1948.

The trial statistician, Austin Bradford Hill, utilised a novel type of randomisation that we would now refer to as a stratified block randomisation. This was in place of the older, and inferior, alternating randomisation that was more commonly used previously [2].

Although they were primarily motivated to ensure any selection biases were minimised, the use of randomisation is now universally recognised as a pillar of sound clinical trial design.

Blinding is typically concerned with ensuring treatment allocations are concealed either for just the patient (single blind) or the physician and patient (double blind), with the aim to minimise any conscious or unconscious biases is the reporting of responses. A control arm is the use of a group of patients who do not receive the experimental drug regimen. These patients may receive a placebo or standard of care, and serve an essential purpose of providing a reference group from which to benchmark the results from the experiment drug regimen. The streptomycin trial utilised a control group, and these patients received the normal standard of care. A rudimentary (by modern standards) blinding effort was used insofar as the control group were unaware they were the control group for the streptomycin group of patients! The use of double blind trials is a feature of many modern clinical trials, utilising further methods such as "double dummy" regimens to mask any differences between the treatment arms. Control arms are, rightly, ubiquitous in modern RCTs, and few would disagree that standard of care type reference arms provide a solid basis for determining the comparative effectiveness of novel experimental drug regimens.

Finally, the trial results showed that 7% (4/55) of the streptomycin treated patients and 27% (14/52) of the control patients died within 6 months, a difference they reported as:

> *"...statistically significant; the probability of it occurring by chance is less than one in a hundred."*

Here we see the adoption of the Fisher/Pearson null hypothesis significance testing (NHST) approach to the interpretation of the results. Although I will subsequently argue that we should view all clinical trial results from an estimation perspective rather than a NHST perspective, the use of the NHST approach by Hill looked to inject statistical rigour to the interpretation of the observed treatment differences, and hence must be applauded.

This rather long discussion of the 1948 streptomycin trial is included because it both serves to highlight many of the successful aspects on modern clinical trials, but also because this type of trial design has inadvertently led to some of the failings I see in modern drug development. Before embarking on the details on these failings, we should also briefly introduce how RCTs fit into the broader clinical drug development programs we see today. The development of a new drug will involves conducting a series of RCTs, and these are broadly grouped into the three phases of drug development:

- Phase 1 trials are precisely controlled trials where small cohorts of individuals receiving a single dose or multiple doses of the drug. These include the so-called first in man trials (FIM), single and multiple assessing doses trials (SAD and MAD), along with additional clinical pharmacology trials designed to answer essential drug development questions (e.g. renal/hepatic impairment, drug-drug interactions, bioequivalence, QTc trials etc.). Phase 1 trials are normally conducted in healthy individuals, with oncology being a notable exception. Treatment durations are typically days to weeks.

- Phase 2 trials are larger trials in patients. These are typically called "dose ranging" or "dose finding" trials, as they investigate a range of dosing regimens; they may also include a placebo and/or an active control arm. Treatment durations are typically week to months. The goal/design is typically stated as to quantify the D-R relationships (more on this later!).

- Phase 3 trials are the largest trials, again in patients. In these trials it is common that only 1-2 doses are compared to placebo / standard of care. Treatment durations are typically months to years. These are often called pivotal trials, since their results will be used to support the approval for the dosing regimens investigated. The goal/design is often based on achieving a statistically significant difference in favour of the new drug over placebo/standard of care for a primary efficacy endpoint.

For the interested reader, this book chapter provides additional background on the history of the drug development process and the role of statistics/statisticians (Statistics and the Drug Development Process). Importantly, it charts how we have continually improved drug development for the better.

We now have sufficient background to discuss the 1948 streptomycin RCT trial in more detail, before discussing the **bad** in the design and analysis of many modern RCTs and drug development programs.

Back in 1948 I am sure there was abundant quackery in the reported benefits of experimental drugs, along with a plethora of anecdotal "evidence" based on limited data from either individual patients or small groups of patients. The streptomycin trial therefore represented an important shift in how new experimental drug regimens were assessed, one centered on good experimental design principles and robust statistical analysis. The design and language used to describe the trial have the "fingerprints" of the foundational work from the Rothamsted Experimental Station on the design and analysis of agricultural field trials from the 1920s onwards, of which Ronald Aylmer Fisher was a leading figure. If you are not familiar with the history of modern statistical analysis, much can be traced back to these early pioneers and innovators. At the time, their "experimental unit" was typically a field or animal, and "yield" was often the key response of interest. Randomisation, Latin square designs, balanced incomplete block designs, factorial designs and hierarchical/nested designs were central components to experimental design, with corresponding analyses based on describing sources of variation (e.g. ANOVA), P values and statistical significance. Their goal was often to determine what combinations of agricultural varieties/practices and quantities of agrochemicals (fertilizers, insecticides etc.) generated, **on average**, the best yield of a crop. Minimising "variation" between experimental units was seen as an invaluable aid in distinguishing real treatment effects from random field-to-field or animal-to-animal variation.

Farmers, at least in the 1920s, were not interesting in the specific results of any individual field, but rather in obtaining the best **average** yield across all fields. It would be unlikely that any farmer would have had the time or energy to consider "personalised" treatments for each field; rather all fields should receive the **same basic treatment**. For example, fields generate some yield without agrochemicals, but with agrochemicals this could be improved; they were turning a good outcome into a great outcome. In addition, treatments needed to be **simple** to ensure they could easily be rolled out to the farmers.

> **!** Important
>
> In **agriculture**, the focus on **average outcomes** across groups of **fields** with the **same treatment** that is **simple** to use **makes sense**.
> In **drug development**, the focus on **average outcomes** across groups of **patients** with the **same treatment** that is **simple** to use **does not make sense.** Why? Because we must **care** about individual patient outcomes. Some flexiblility with dosing is a small price to pay for better patient outcomes.

We can therefore list 6 key considerations when we plan the design and analysis of modern, patient-centric, RCTs:

- We need to look beyond average outcomes with the same, simple drug regimen.

- We must care about (and hence optimise) individual patient outcomes; **patients are not fields**!

- Patients will normally need different doses.

- Drug development needs different/better designs and analyses than basic agricultural designs and analyses. Simple does not always suffice.

- Patient heterogeneity is not a "nuisance" variation; it is real and important.

- Estimation, and not significance testing, is the goal of the analysis.

Regarding patient heterogeneity, the 1948 streptomycin report specifically commented on "eliminate…variations" regarding the inclusion/exclusion of patients. In modern drug development, we should see all patient groups who may ultimately receive the approved drug as candidates for recruitment in the supporting drug trials. Thus inclusion/exclusion criteria should not be used to "reduce variation" in the patient population; this is a clear distinction between agricultural trials were homogeneous fields may be appropriate, and drug development for patients, where we seek the best dosing regimens for each one of our heterogeneous patients (we want solutions for **all** patients). The idea that drug regimens should be approved for a wide range of patients based on trial results from a select, narrow subgroup of the wider patient population is both incoherent and unacceptable. For example, if an 80-year-old patient on multiple medications would be considered a candidate for the drug **post-approval**, then should we not have obtained data from similar patients within our well-controlled clinical trials **pre-approval**, where a multitude of efficacy and safety parameters would have been collected and analysed? **Ignorance post-approval by sidestepping such patients pre-approval is wrong.** This is not to say that inclusion/exclusion criteria cannot be more restrictive earlier in the development of a drug, when both efficacy and safety data is limited. However at some stage pre-approval we must collect data in a patient population that is truly representative of those who will ultimately receive the drug post-approval.

Throughout this book, the emphasis is always one of **estimation**, not of **significance testing**. The significance testing approach to design/analysis is concerned with being able to reject statements such as "the drug effect is not zero," whereas the estimation approach is concerned with **quantifying** how measures of efficacy and safety/tolerability change as a function of the dosing regimen. **Estimation is concerned with accurately and precisely answering the "how much" question**. With significance testing, the idea of collapsing a distribution of an estimate of interest (like a treatment difference) to a binary yes/no based on whether the P value is less than 0.05 is unhelpful. Thus the notion that two placebo controlled phase 3 trials with P<0.05 should be acceptable for regulatory approval is laughable. The drug company has only successfully demonstrated (twice) that the drug effect is not zero for a single primary efficacy endpoint! **This is no "gold standard"**. For any meaningful assessments of the benefits and harms, we need to precisely quantify how multiple efficacy and safety/tolerability endpoints change as a function of the dosing regimen (i.e. across a range of doses if we wish to determine a suitable dose range for approval). In cases where only a single dose level is investigated in phase 3, there is clearly no opportunity for either the sponsor or the regulator to determine whether the dosing regimen is, in any way, optimal. Such weak data also inhibit any understanding of how steep or flat the actual D-E-R relationships are, and hence we are unable

to quantify the risks to patients who have particularly high drug exposures, or inadvertently take a higher dose than planned. Frankly, this is just poor drug development.

We have covered a number of **good** features of modern RCTs such as randomisation, blinding and control groups. We have also mentioned some of the **bad** features, such as focusing on average outcomes rather than individual outcomes, the fixation on simple dosing regimens, and the overuse of statistical testing and P values (rather than focussing on the precise estimation of the effects for different dosing regimens). In the final section I will briefly cover some of the **ugly** features. The list could be longer, but I wanted to draw attention to 3 particular failures that I find unacceptable. As an ethical scientist, these do not sit well for me, and I hope you would agree.

Firstly, is it use of "off-label" dose regimens, both for adults and children. The routine use of off-label doses is the proverbial "elephant in the (regulator's) room," and contravenes the fundamental principles of evidence-based medicine. For example, at the European Medicines Agency Dose Finding Workshop meeting in December 2014, one physician stated how he used most antipsychotics at doses substantially different from the approved doses. I expected the room to explode into discussion. Instead, there was no discussion, and the report simply noted for this presentation:

> *"This means that the benefit/risk balance of the 'real' doses has never been subject to regulatory scrutiny."*

Is it acceptable that we routinely expose individual patients to doses beyond those studied and documented? Is this "wild west" approach to dosing, in which "anything goes," acceptable? **In short, drug companies and regulators cannot sit in ivory towers with pieces of paper saying how the drug should be used, whilst allowing patients to be given untested dosing regimens by well meaning but misguided physicians**. If novel dosing regimens are to be tested, surey these must be investigated in appropriate RCT that have obtained ethics approval and informed consent.

Secondly, is the overuse of placebo controlled trials in therapeutic areas where numerous, effective treatments exist. For example in type 2 diabetes, patients enter clinical trials with significant hyperglycemia. Despite over 40 drugs being approved for type 2 diabetes, we continue to randomise patients to placebo. Given sitagliptin is a modestly effective drug with an excellent safety profile, surely we should replace placebo with sitagliptin (or any other approved treatment)? Sitagliptin is one drug in a large model-based meta-analysis I have conducted [3] including over 300 trials, of which over 40 trials included over 12000 sitagliptin treated patients; we know how effective sitagliptin is, and should use this information instead of continuing to randomise patients to placebo (see the diabetes MBMA here (*tip: if you follow any link, use the "back" button in your browser to return to where you were*)). In psoriasis, again it is common to randomise patients to placebo, despite over 15 drugs being approved. Psoriasis can be an awful, debilitating disease, and patients with psoriasis who enter clinical trials are in real need of effective treatment; many are in daily pain. We should not consider it acceptable to expect patients to endure 4 months of ineffective placebo treatment when

a whole host of effective treatments are available. Whilst we may find it "convenient" and "easy" to continue with placebo arms, I would simply ask you this "Would you be happy to see you son or daughter receive placebo in such a psoriasis trial, when a plethora of effective treatments exist?"

The final **ugly** reality of modern drug development is perhaps less heinous than those above, but something that continues to infuriate me! When we run any clinical trial in patients, the trial must be ethical; it must be able to generate meaningful and useful data to answer the questions that it seeks to address. Poor experiments in humans should not be tolerated. So let's talk about phase 2 "dose-finding" trials. If I had a dollar for every time I read the objectives of a trial as:

"To assess the dose-response relationship of [DRUG] on [ENDPOINT] in subjects with [INDICATION]"

Only to then see a terribly designed and analysed D-R trial, I would be many hundreds, perhaps thousands, of dollars richer.

Since this is so important, the whole next chapter is devoted to this topic!

---

5 Dose-Response Trials; A Brief History And Overview Of Current Practices

At the end of this chapter, the reader will understand:

- How the 1994 ICH E4 guideline "Dose-Response Information to Support Drug Registration" has shaped modern phase 2 D-R trial designs.

- How ICH E4 rightly distinguishes between **Population** (average) D-E-R relationships determined from **Parallel**, fixed-dose, D-R trials and **Individual** D-E-R relationships determined from **Individual Titration** D-R trials.

- How the **simplicity** of the analysis, and **not** optimizing individual patient outcomes, led ICH E4 and the FDA to wrongly promote **Parallel**, fixed-dose, D-R trials over **Individual Titration** D-R trials (since the latter required "careful analysis"!).

- How most modern phase 2 D-R trials are poorly designed, poorly analysed, or both.

- If we wish to refocus drug development to use dose to maximise individual patient outcomes, we need **Individual Titration** D-R trials that can determine **Individual** D-R relationships.

On the topic of how to design D-E-R trials so that we learn how best to dose each and every patient, we will start with a (mis) quote from the journalist H L Mencken:

"For every complex problem there is an answer that is clear, simple, and wrong"

Most of the designs and analyses of D-E-R trials are indeed clear, simple, and wrong. Let me explain why.

To understand modern D-E-R trials, two important events/observations from the past need to be discussed. Firstly the critical role played by the 1994 ICH E4 guideline "Dose Response Information to Support Drug Registration" [4]. Although this document in nearly 30 years old, the authors made many sensible observations; a few are included and discussed below.

On the issue of selecting a starting dose:

> *"What is most helpful in choosing the starting dose of a drug is knowing the shape and location of the population (group) average dose-response curve for both desirable and undesirable effects. Selection of dose is best based on that information, together with a judgement about the relative importance of desirable and undesirable effects."*

Here the authors recognized that the judicious selection of the starting (initial) dose can be guided by understanding the shape of the (population) D-E-R for both efficacy and safety/tolerability (without this information, there is indeed no scientific basis for the starting/initial dose).

On the issue of how to titrate the dose for a patient:

> *"In adjusting the dose in an individual patient after observing the response to an initial dose, what would be most helpful is knowledge of the shape of individual dose-response curves, which is usually not the same as the population (group) average dose-response curve. Study designs that allow estimation of individual dose-response curves could therefore be useful in guiding titration, although experience with such designs and their analysis is very limited."*

This astutely acknowledges that patients follow their **own** individual D-R relationship, and that this is **different** to the population D-R relationship.

On the issue of the design of parallel, fixed-dose, D-R trials:

> *"The parallel dose-response study gives group mean (population-average) dose responses, not the distribution or shape of individual dose-response curves."*

> *"It is all too common to discover, at the end of a parallel dose-response study, that all doses were too high (on the plateau of the dose-response curve), or that doses did not go high enough. A formally planned interim analysis (or other multi-stage design) might detect such a problem and allow study of the proper dose range."*

Thus the authors understood that the outcomes from such trials only yield the population (average) D-R; we do not learn individual D-R relationships. In addition, they mention the value of an interim analysis/adaptive design to ensure the right dose range is actually explored in such trials (discussed further in the chapter on **adaptive randomisation**, see Chapter 19).

Under the heading "**Parallel dose-response**" the guideline states (with my emphasis in **bold**):

> *"Randomization to several fixed dose groups (the randomized parallel dose-response study) **is simple in concept** and is a design that has had extensive use and considerable success."*

Under the ominous heading of "**Problems with Titration Designs**" perhaps the most relevant sentence to explain why we are where we are today is the following (with my emphasis in **bold**):

*"A study design widely used to demonstrate effectiveness utilizes dose titration to some effectiveness or safety endpoint. Such titration designs, **without careful analysis**, are usually not informative about dose-response relationships. In many studies there is a tendency to spontaneous improvement over time that is not easily distinguishable from an increased response to higher doses or cumulative drug exposure. This leads to a tendency to choose, as a recommended dose, the highest dose used in such studies that was reasonably well tolerated."*

In 2018 a highly experienced regulator, Robert Temple (Deputy Center Director for Clinical Science at the FDA), gave a presentation on the Design of Clinical Trials.

It included this text regarding D-R:

"Goal: Define D/R curve for benefits and risks

Until early 1980's, most trials with more than one dose titrated the dose, generally to some endpoint. This meant:

1. *The group on any given dose was not chosen randomly*

2. *Time and dose were confounded; secular trend would look like response to dose. Particularly useless for safety*

In 1980's, FDA promoted the randomized, parallel, fixed dose, dose-response study, identified as the standard in ICH E4 guidance."

Thus the above texts provide us with an explanation of why parallel, fixed-dose designs for D-R trials where (incorrectly) deemed preferable to titration based designs (for brevity, these will be referred to a **Parallel** and **Individual Titration** below).

> **!** Important
>
> **Parallel**, fixed-dose designs for D-R trials were promoted as **superior** to **Individual Titration** trials because they are just "simpler" to **analyse**. However they are **not** best for patients!

Both the ICH E4 text and that from Robert Temple rightly state that a simple analysis of titration trials that ignores the (partial) confounding of dose and time would be flawed; however this absolute does not mean they titration trials are unanalyzable – it just means exactly what ICH E4 states, a "simple analysis" does not suffice! Indeed, the difference between fixed-dose regimens and placebo in a **Parallel** trial **also** changes as a function of time (duration of treatment), but no one would suggest we cannot describe/model how the dose effects evolve over time (relative to the placebo response, which also changes with time!). Furthermore, whilst the dose may remain the same over time, the drug concentrations achieved with repeated doses typically increase over time, so a "fixed-dose input" is "fixed" in name only, as the system (the patient) actually experiences a "changing concentration input" over time with repeat dosing

(if this point is unclear, later chapters will introduce the key PK principles that underpin and explain this observation).

In addition, there are both errors and significant omissions in ICH E4. Under "Guidance and Advice", it states:

> *"A widely used, successful and acceptable design, but not the only study design for obtaining population average dose-response data, is the parallel, randomized dose response study with three or more dosage levels, one of which may be zero (placebo). From such a trial, if dose levels are well chosen, the relationship of drug dosage, or drug concentration, to clinical beneficial or undesirable effects can be defined."*

The above suggests that only 3 dose levels (such as placebo, 10 mg and 20 mg) would be a "successful and acceptable design" to define both benefits and harms. This is incorrect. To fit an appropriate D-R model, you need a minimum of 4 dose levels, and these need to be well spaced over the "right" part of the D-R relationship (see Chapter xxx). To be clear, a design such as Placebo, 10 mg and 20 mg is incapable of adequately describing any D-R relationship.

ICH E4 also has a glaring omission; it fails to suggest how to actually "link" doses together using a suitable D-R model(s). **No D-R models are discussed or considered; the actual analysis of D-R trials is wholly absent**. How can any D-R relationship be determined without a D-R model? Clearly one can "join the dots" of the mean responses at each dose (like in the Placebo, 10 mg, and 20 mg example), but this is an awful and painfully unscientific strategy (this will be illustrated later in Chapter 11). Since ICH E4 is still an active guideline, I feel it is important to shine a light on these major errors and omissions. Good D-R trial design requires the investigation of well-spaced dose levels over a wide dose range in sufficiently large sample sizes. The analysis then combined all data together using a suitable D-R model. Thus we need both the right design and the right analysis, and that requires a D-R model!

Finally, it is important to note that the criticism of parallel group dose ranging trials herein is not new. In a 1989 article entitled "Study Designs for Dose Ranging", Sheiner, Beale and Sambol [5] wrote:

> *"We believe one must begin with a parametric model for patient-specific dose-response curves. Knowledge of the distribution of these curves in a population provides a basis for choice of an initial dose (e.g., the dose that achieves a given response in a given fraction of patients) and, after observation of response to an initial dose, for choice of an incremental dose for a specific patient (by use of Bayes rule). The current parallel-dose design can provide only poor information about the distribution of dose-response curves, biased estimates of the typical curve, and little information on interpatient variability"*

In their discussion, they added:

> *"The dose-escalation design imitates clinical practice and was a popular design for dose-ranging studies in the United States (it continues to be so in Europe) until replaced by the parallel-dose design. It is instructive to examine why the latter appeared preferable. In brief, we believe design and analysis flaws, which are obviated to a large extent by the approach described in this article, were responsible. Apparently, rather than analyze and correct the flaws, researchers chose to abandon the design. In our opinion, this was a case of throwing out the baby with the bathwater."*

I fully agree.

Returning to our primary comparison around **Parallel** and **Individual Titration** designs, we need to be very precise in the comparison between these two D-R designs.

Parallel and Individual Titration D-R designs are different, the data acquired is different, the analysis is different, and the focus of inference is different (population D-R or individual DR). The questions they address have both similarities and differences.

At the highest and simplest level we can compare the two designs based on the final outcomes they achieve (versus placebo or an active control). For example:

- **Parallel -** if all patients receive the same **fixed-dose regimen**, what are their outcomes after 6 months?

- **Individual Titration -** if all patients use the same **dose titration algorithm**, what are their outcomes after 6 months?

Thus we see our first similarity. Both designs generate a set of individual patient outcomes (for both efficacy and safety) that can easily be compared. However with **Parallel**, we are forced to maintain the exact same dose throughout, **irrespective of how well or poor the intermediate outcomes are for each patient;** despite on-treatment data/evidence, we deliberately aim to not change the dose! As such, I see such fixed-dose regimens as appealing to fatalism – we are (apparently) resigned to the fact that the benefits and harms caused by the dose to the patient are unavoidable and inevitable; the patient **must** experience the **full** frequency and severity of tolerability/safety effects that has been assigned to them; this is not OK. Our need for "simple" data is being placed above our need for better patient outcomes; this is wrong. With **Individual Titration**, we are guided to change the dose to maximize the benefits and minimize the harms caused by the dose for each patient. In contrast to the **Parallel** patients, the dose for each patient may be adjusted based on accruing observed intermediate outcomes, and hence we have the opportunity to use dose to **improve** their final outcomes.

> **❗ Important**
>
> In a **Drug Development For Patients** world, if we have the opportunity to **improve** individual patient outcomes by **changing** their dose, are we not compelled to do so? **I strongly believe we are. Do you?**

For well designed **Parallel** D-R trials, we get the necessary information to construct the **Population** D-R that, in some cases, will be reasonable. For example, we may get 100 patients for placebo and 5 different dose levels (600 patients in total). Thus the resulting D-R modelling will determine the population (average) effect at each dose under the presumption that we will give all patients that dose (i.e. with no titration). For **Individual Titration,** we would get 100 placebo patients and 500 individual D-R curves. Depending on the design used (e.g. forced titration), patients will generate data across their own dose range, and the resulting D-R modelling would determine **Individual** D-R relationships. Clearly care needs to be taken in both the design and analysis of these trials, but rest assured, technically we can do this!

It is worth stressing that population D-R effects are only accurate if we specifically **prohibit** any dose titration (from a "do no harm" perspective, why would we want to prohibit dose titrations?). For example, if a population D-R analysis determined that a dose of 100 mg would result in an average heart rate increase of 10 bpm, we may not expect the same average 10 bpm increase at 100 mg if this dose was towards the higher end of a dose titration range (e.g. 10-100 mg). This is because some patients (e.g. those with higher than average drug concentrations and/or more sensitive to the drug) may be adequately treated at lower doses, and hence never reach 100 mg. Thus from a titration perspective, we would be more interested in the heart rate effect at 100 mg **only** for those patients who are actually titrated from, say, 50 mg to 100 mg (that is, the cohort of patients who were inadequately treated at 50 mg). Population D-R does not answer this important question; rather it only ever addressed the "one-size-fits-all" dose effect when any form of titration is specifically prohibited.

With **Parallel** D-R trials, we answer a simple question, but in doing so place a paralyzing inability to adjust the dose for the patient, something we know we need to do! For anesthetists it would also be "simpler" to initiate the infusions/flow of anesthetic agents and then go home, but there is a good reason why they do not. They care about, and seek to optimize, individual patient outcomes.

When we weigh up **Parallel** versus **Individual Titration** D-R trials, we may ask "Do we want an easy design/analysis that answers a less important question, or a design/analysis that can answer the right question?"

In his book entitled "Dose Finding in Drug Development", the author Naitee Ting wrote:

> *"There are some advantages of a titration design. For example, a study with this design will allow a patient to be treated at the optimum dose for the patient; this dose*

allocation feature reflects the actual medical practice. However, the disadvantage of a titration design is the difficulty in data analysis."

I think this "**screams**" the choice we must make; should we choose the path that allows each patient to be treated at their optimal dose, or should we choose the path that avoids a difficult analysis? **Do we care more about having a simple analysis than our patients?!** Perhaps to state the problem more pointedly, should we allow a patient with cancer to die on their fixed-dose regimen, because we cannot be bothered to monitor markers of efficacy (e.g. tumour growth) and hence consider a dose change for them? If your partner/child were in such a trial, what design would you choose, and why?

In summary, the above text has sought to illustrate why **Parallel** group D-R trials were originally seen by some in the 1980s/1990s as superior to **Individual Titration** D-R trials, however the primary motivation was based on the simplicity of the analysis, and not on obtaining the best patient outcomes. **Thus although ICH E4 is still sound in many regards, it is also now painfully outdated and desperately needs updating**. It fails to show the significant weaknesses with **Parallel** D-R trials, and the real value to patients of **Individual Titration** D-R trials. If our goal is to have simple trials and "one-size-fits-all" doses, then **Parallel** is OK**.** If our goal is to obtain the best outcome for each and every patient via informed **Personalised Dosing,** we need well designed **Individual Titration** D-R trials.

Given that above history, it is perhaps not surprising to know that the current landscape of phase 2 dose-ranging trials is dominated by **Parallel**, fixed-dose, D-R trial designs. Unfortunately not only are these designs ubiquitous, the standard of their designs and analyses are frequently very poor.

To illustrate this point, I searched "Clintrials.org" for completed phase 2 trials with the statistical analysis plan available (so I could find the more technical information). Table 5.1 below shows some details of the first 5 trials I found (to be more representative of the industry as a whole, I did select trials from larger pharma companies, since you may expect these companies to have sufficient resources to design these trials well).

These trials were generally designed in the last 5-7 years, and look representative of the (weak) designs and analysis of "dose-ranging" trials I often see.

Table 5.1: An illustration of recent dose-response trials

―――――――――

―――――――――

Trial Number	NCT03375203	NCT02966834	NCT02973321	NCT03233230	NCT02447302
Company	J&J/ Janssen	GSK	Sanofi	EMD /Merck	Pfizer
Drug	seltorexant	linerixibat	SAR425899	evobrutinib	etrasimod
Indication	insomnia	primary biliary cholangitis (PBC)	type 2 diabetes	rheumatoid arthritis	ulcerative colitis
Objective	To assess the dose-response of 3 doses of JNJ-42847922 (5, 10, and 20 mg) compared to placebo on an objective measure of sleep onset in subjects with insomnia disorder.	To investigate the dose response of oral GSK2330672 on itch in PBC patients with moderate to severe pruritus at Baseline.	The primary objective of this study is to assess the dose-response relationship of SAR425899 versus placebo in terms of glycemic control as measured by the change in glycosylated hemoglobin (HbA1c) from baseline to Week 26.	To evaluate the efficacy and dose response of 12 weeks of treatment with evobrutinib compared with placebo in subjects with rheumatoid arthritis.	APD334-003 is a phase 2, proof-of-concept and dose ranging clinical study designed to test the safety and efficacy of APD334 in patients with moderately to severely active ulcerative colitis.
Treatments /Doses	Placebo, 5mg, 10mg, 20mg, Zolpidem	Placebo 20 mg 90 mg 180 mg 90 mg bid (20mg dropped for 40 mg bid) (adaptive trial)	Placebo, 0.12 mg, 0.16 mg, 0.20 mg, liraglutide	Placebo 25mg, 75mg, 50mg bid	Placebo, 1mg 2mg
N (total)	363	136	296	390	156
Sample Size Justification	Power	Using a simulated Emax D-R model, they state "A sample size of approximately 100 eligible participants allows sufficient precision to provide a minimally detectable effect of at least 2 points"	Power	Power	Power
Analysis	ANCOVA estimates => MCP-MOD	ANCOVA => sequential model fitting process (see next figure)	Trend test	ANCOVA (MCP-MOD was the 6th supportive analysis)	ANCOVA
Safety Analysis	No	No	No	No	No
Dose Individualisation	No	No	No	No	No

Relative to the objective, to understand the dose-response relationship, these trials are universally weak. The designs have very few dose levels across very narrow doses ranges (4/5 trials). The sample sizes are based on power statements (4/5 trials), which means they are justifying the sample size on the ability of the design to reject the null hypothesis of "no drug effect". This is completely different to understanding the dose-response! **Being able to conclude at the end of the trial that "the drug does something", is not the same as being able to accurately and precisely quantify the D-R; the latter requires appropriate designs and greater sample sizes.** The analysis methods are poor, either based on pairwise comparisons (i.e. ANCOVA), trend tests or the better (but still weak) MCP-MOD (more on this later). **In addition, none of these trials reported any planned D-R analysis of safety/tolerability endpoints in the statistical analysis plan (although there may well be a pharmacometrics analysis plan with such details).** A critical objective of drug development is to understand how safety/tolerability change as a function

of the dose, yet here it is unclear if any D-R analysis of the safety/tolerability endpoints has been prospectively planned, or if the design is suitable for these crucial analyses. Without such analyses key decisions, such as phase 3 doses to consider, may be forced to rely on just the observed point estimates for each dose (the limitations of which will be discussed in later chapters). **Finally, none of these trials considered dose-individualisation (i.e. Individual Titration D-R trials); all were fixed dosing regimens.**

The GSK trial was less disappointing, with a 9-fold dose range, and a sample size justified based on something other than a simple power statement. However this trial came with a flow chart for the analysis. This is shown in the Figure 5.1 below, and is a clear example of poor D-R modelling.
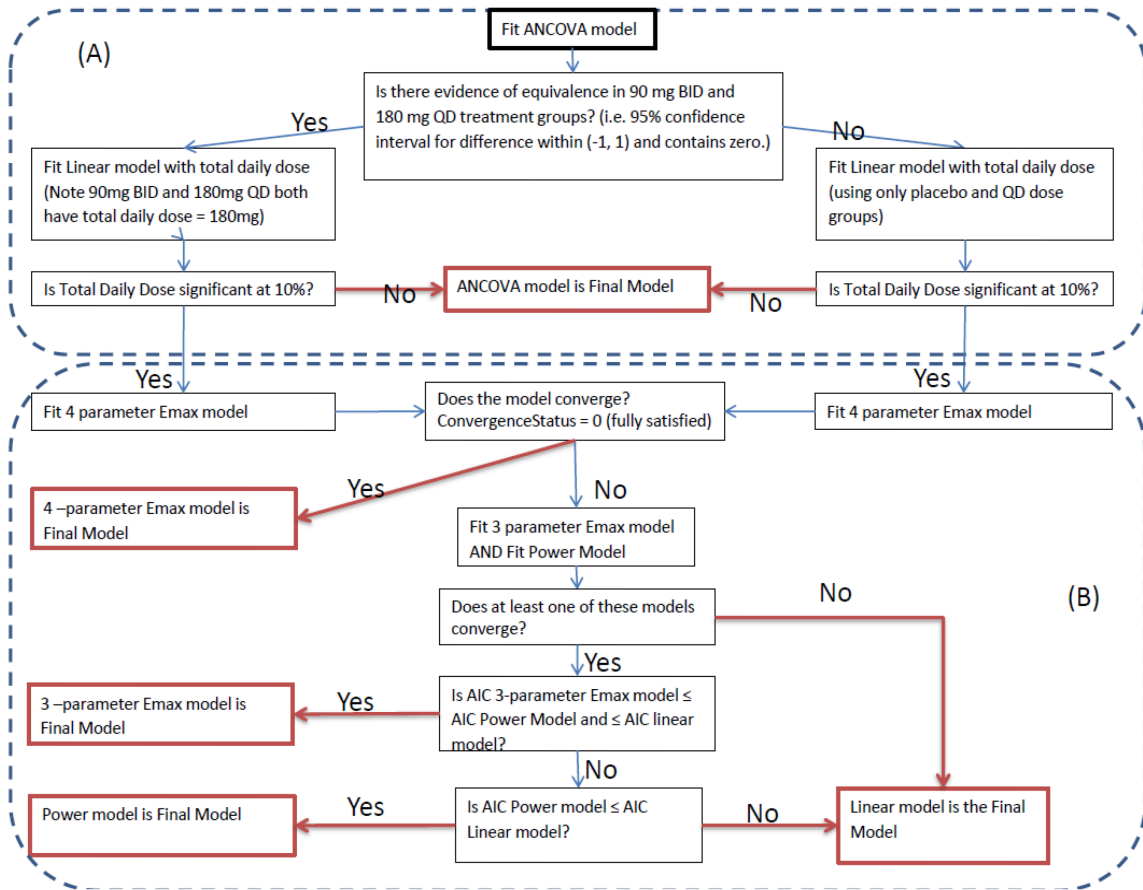

Figure 5.1: GSK Statistical Analysis Plan for linerixibat phase 2 trial - D-R modelling plan

The diagram is a mix of questionable logic, significance testing and poor model selection criteria; **perhaps my reaction is extreme , but I see this D-R analysis being brutally butchered before my eyes!** Firstly, the 90 mg bid versus 180 mg qd doses are tested as to

whether we can reject the hypothesis that they are equivalent. Deciding whether to progress a once or twice daily regimen is not trivial, since the differences between these two regimens will be generally small; you need very large trials to quantity these differences. Thus this first step can only lead to pooling the two regimens (hence assuming once and twice daily are identical, which they are not) or, worse still, concluding the twice daily regimen is **markedly** better than the once daily regimen, which in itself would be a spurious finding (that is, you will only conclude the twice daily is significantly better in the few occasions where, by chance, the observed difference was unusually large). The analysis does then look to fit the very sensible 4 parameter (sigmoidal) Emax model, but then relies on the software to decide if this model should be accepted! If a sensible D-R model fails to "converge", you must first ask "Why?" This will require both the data and the software fitting information output to be reviewed, and then to determine how best to proceed (perhaps the model simply needs to be reparameterised). Instead, the analysis blindly follows a sequential set of ever more naive models, potentially leaving us with the (never appropriate) linear D-R model. From an "easy of upfront programming" perspective, the flow diagram is wonderful. Unfortunately from an "accurately and precisely quantify of the D-R relationship" perspective, it is not sound.

Please understand I have no issue with any of the companies above, and the very small sample of trials above may not be a fair reflection of how any of these companies typically approach dose-response trial designs (e.g. a friend mentioned that the Pfizer trial is a phase 2A trial, with wider dose ranges studied in phase 2B). In addition, I could equally have commented on other major companies/trials, including Roche/Genentech (Fenebrutinib (NCT02833350)), Novo Nordisk (semaglutide (NCT02453711)), Abbvie (tilavonemab (NCT02880956)), Novartis (LCI699 (NCT00758524)) and Lilly (tirzepatide (NCT03131687)). All these dose-ranging trials are limited in either their design, their analysis, or both. The key point here is **not** to "bash" any company, but rather to show that weak D-R trials are quite common across industry, and not just limited to a few "rogue" companies.

There are also examples of much better D-R trial designs that considered very wide dose ranges (as advocated herein). For example placebo and a 40-fold dose range (Pfizer trial NCT03985293), a 60-fold dose range (Novartis trial NCT03100058), and a 16-fold dose range (GSK trial NCT00950807). This is very important, because knowing if lower doses may do as well as much higher doses for efficacy, but much better for safety/tolerability, is crucial to finding the right dose range for patients. For example in trial NCT03926611 Novartis observed similar efficacy of remibrutinib versus placebo across the dose range, from 10 mg qd to 100 mg bid (a 20-fold dose difference!). Without these lower doses, perhaps the 100 mg bid dose would have been progressed into phase 3, even though much lower doses appear similarly effective. **It is also worth mentioning another excellent trial at the right end of the spectrum, one by Lilly** [6]. Stage 1 of the dulaglutide phase 2/3 adaptive trial (AWARD-5 (NCT00734474)) investigated a 12-fold range of dulaglutide doses simultaneously across 2 efficacy endpoints and 2 safety endpoints using a Bayesian adaptive, D-R design. They sought to determine the two doses with maximum utility (the "trade off" between efficacy and safety) in stage 1, before selecting these doses for stage 2 (the phase 3 part of this seamless trial). Although there are many ways this trial could have been even better, it was clear that the

team prospectively evaluated (via simulation) how to design the trial to combine D-R models for efficacy and safety. Stage 1 of this design is particular good, and reflective of the sound work of Donald Berry and colleagues; stage 1 is actually similar to one type of drug development strategy that will be discussed in Chapter xxx, and hence is worth reading.

In contrast to the wide dose ranges and adaptive trial designs discussed above, most oncology "dose-finding" trials are particular weak, as often only 1-2 (very high) doses are actually considered in phase 2/3. Given the myriad of serious safety/tolerability problems experienced by patients with many oncology dosing regimens, the lack of well-designed and analysed D-R trials in oncology is particularly disappointing. **Simply exposing patients to very high doses is neither scientifically sound nor ethical drug development.**

By the end of this book, I hope it will be clear why we generally need to study very wide dose ranges in sufficiently large sample sizes. We need our trials/programs to accurately and precisely quantify the D-R relationships for efficacy and safety/tolerability. In addition, most trials only consider fixed-dose regimens, with the expectation that a single "one-size-fits-all" dose will be appropriate for all patients. The idea that patients could achieve **better** outcomes with individualised dosing (i.e. personalised dosing) is often regrettably absent from these trials; I hope this will change. **Remember, patients are not fields!**

This section has hopefully given a short introduction to the history of D-R trial designs, and how ICH E4 has unfortunately led to **Parallel,** fixed-dose, D-R trials to determine **Population** D-R relationships being used because they are "simple". In addition, a brief review showed that some of these trials are of a low standard. Finally, if we wish to refocus drug development to use dose to maximise individual patient outcomes, we need **Individual Titration** D-R trials that can determine **Individual** D-R relationships; Sheiner, Beale and Sambol were right .

6 The Science; Why We Must Care About Dose, Pharmacokinetics, Pharmacodynamics And Utility

At the end of this chapter, the reader will understand:

- The basic concepts and interrelationships between dose, Pharmacokinetics (PK), Pharmacodynamics (PD) and Utility.

- Why inter-individual variability (IIV) in PK and PD is the primary reason why individuals need different doses.

- That IIV in PK yields a range of exposures (concentrations) and IIV in PD yields a range of responses (across individuals receiving the same dose).

- For every endpoint, each patient has their **own** individual D-E-R relationships.

- That dose is just a very crude mechanism to attempt to deliver sufficient drug to the site of action to illicit the desired PD responses.

- Why we should **always expect** to need to change the dose to achieve the best responses in all patients.

- Drugs that are suitable for "one size fits all" dosing are like diamonds; they are very rare.

For motor racing drivers, they want a car that gives them the best outcome; they win the race. To enable this, the driver needs teams of expert engineers to design, develop, build, and optimise the car. The fastest car will involve multiple experts (aerodynamicists, material scientists, engine designers etc.) to continually optimise their "product" (the car) through the use of well design experiments and modelling and simulation. In drug development, the patient is our driver; they rightly just care about their outcomes. However to use our drug as well as possible, we need drug developers and regulators who have the same "engineering"

mindset to use science and well controlled experiments (trials) to make our "product" as good as it can be. Determining the best way to dose each and every patient is, like a motor racing car, not "simple", and we should not pretend it is. So let us consider the science.

Figure 6.1 below shows the sequence we aim to understand for all patients, the **individual** D-E-R relationships, and ultimately the utility (benefit-risk) for each patient.

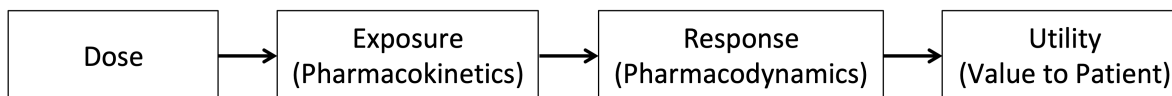| Dose	→	Exposure (Pharmacokinetics)	→	Response (Pharmacodynamics)	→	Utility (Value to Patient)

Figure 6.1: The sequence from dose to utility is **different** for **different patients**.

The term **Exposure** is used herein to broadly refer to measures of the drug concentrations achieved in the body, and the branch of pharmacology dedicated to understanding this is called pharmacokinetics.

In short:

- PK can be defined as what the body does to the drug.

- PD can be defined as what the drug does to the body.

- Utility can be defined as the overall evaluation the patient assigns (collectively) to their individual responses.

There are numerous excellent textbooks and resources for both PK and PD, but herein we will briefly cover the essential details that need to be appreciated if we wish to get the best outcomes for patients through informed clinical trial designs.

A central component of PK is concerned with the absorption, distribution, metabolism and elimination of the drug (ADME). These play a central role in determining the blood plasma concentrations of the drug observed over time. These drug concentrations may be summarised at a particular time point (e.g. C_{trough}, a "trough" drug concentration measured just prior to successive dosing events), or as the maximum drug concentration (C_{max}) or as an average drug concentration (C_{ave}) over the intra-dosing interval. Key concepts in PK include accumulation (when drug concentrations increase over time with successive doses), steady state (accumulation ends, and the drug concentrations reach a dynamic equilibrium with successive doses), and dose-proportionality/non-linearity (whether or not the magnitude of the concentrations change directly with the magnitude of the changes in dose). These basic concepts are shown in Figure 6.2 below.

Here the individual receives weekly dosing for 6 weeks (gray triangles); steady state is achieved after around 3 weeks. Representative values for C_{max} and C_{trough} are highlighted, along with a visual representation of C_{ave}, the average "area under the curve", after the last dose (C_{ave} will always be between C_{max} and C_{trough}).
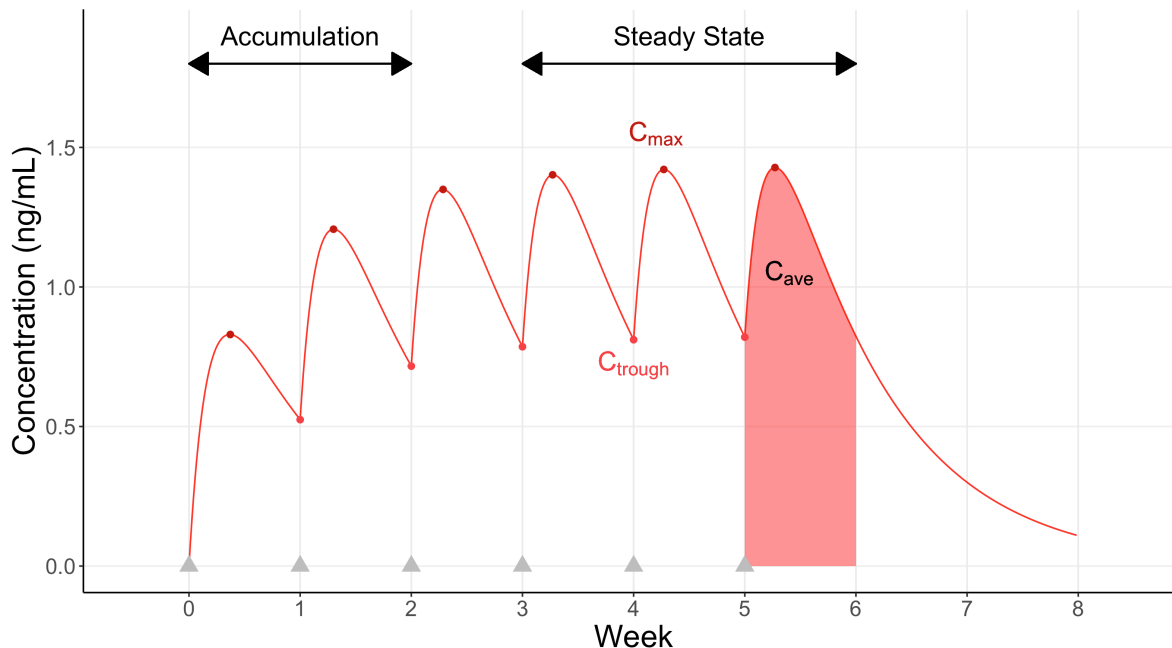
Figure 6.2: Illustration of the PK profile for one individual with repeat dosing.

In general, blood plasma drug concentrations are very important, as this will be the key pathway for the drug to reach the site of action (exceptions are typically due to the route of administration, such as inhaled drugs (e.g. asthma) or topical drugs (e.g. psoriasis)). Additional important PK considerations include drug-drug interactions (whereby taking two drugs together changes the PK of one or both drugs) and the role of active metabolites (when the (parent) drug is metabolised into a molecule that itself can induce a PD response).

PD is concerned with the biologic effects of the drug/dose/concentration, and hence are the responses that matter to the patient. These responses are often labelled as benefits (i.e. efficacy) and harms (i.e. safety/tolerability), and thus include all clinical endpoints that we may measure in a clinical trial.

Although there are cases where the PD effects are observed to change almost instantaneously with the changes in PK (a "direct" effect), it is much more common that PD changes are delayed relative to the PK concentration (an "indirect" effect). For example, the skin of a patient with psoriasis will not immediately change following their first dose, but rather will improve more slowly over the comings days/weeks/months.

The field of PK/PD modelling has developed an extensive range of flexible models to describe simultaneously how both the PK and PD change over time, with the PD changes over time being driven by the PK changes over time (with suitable delayed effects where necessary). In cases where the delay in response is substantial (e.g. weeks or months) relative to the dosing regimen frequency (e.g. daily), it is often reasonable to relate a simple measure of drug

33

exposure, such as the average concentration (C_{ave}), to the PD endpoint, since the fluctuations in concentrations between the dosing events (e.g. within the day) are unlikely to add further clarity beyond that provided by the simpler drug exposure measure. Notable exceptions would be when particularly high concentrations may be more strongly linked to an acute toxicity than the average concentration, in which case C_{max} may be more informative that C_{ave}. The usefulness of C_{trough} is primarily in its simplicity, since it requires only a single PK sample to be taken (unlike both C_{max} and C_{ave}). In cases where there is a very strong correlation between C_{trough}, C_{ave} and C_{max}, it may be acceptable to rely on a single PK sample, but generally it is much wiser to collect multiple PK samples from each patient in the drug development program, since this will allow their full PK profile over time to be predicted, and hence allow a much more accurately understand the interrelationship between the PD effects observed and the PK effects that are driving them.

Utility is a more nuanced topic that will be discussed in more detail in later chapters, but essentially is concerned with how patients actually value the trade-off between the benefits and harms. For example, if an epileptic patient achieved a 50% reduction in seizures with their initial dosing regimen, but experienced 1-4 moderately intensive headaches each month thereafter, would they consider this trade-off worthwhile (a positive utility) or not (a negative utility)? In addition, would they prefer to explore lower/higher doses to find a better dose for them (one with a higher utility)? For the remainder of this chapter, the role of utility will be temporary paused, and we will focus on the first three components, the sequence from dose to PK to PD across our heterogeneous patients.

As Woodcock [7] wisely wrote:

> *"The principal challenge in therapeutics is the variability of human responses to drugs, both for good and for ill."*

If we wish to intelligently dose drugs, I would posit that the most important aspect to understand and account for is the inter-individual variability (IIV) in both PK and PD across patients.

That is:

For every endpoint, each patient has their own individual Dose-Exposure-Response relationships.

This means that every patient will have their own individual Dose-Exposure relationship (IIV in PK) and their own individual Exposure-Response relationship (IIV in PD). Thus any fixed-dose regimen given to a group of patients will yield a wide range of exposures (drug concentrations), and the responses to these drug concentrations will also differ between patients. Figure 6.3 shows one way we can illustrate the relationship between Dose, Exposure and Response for 6 hypothetical patients depending on whether we have a "Fixed Dose", "Fixed Exposure" or "Fixed Response" (these are essentially selected points on the individual D-E-R curves for each of the 6 patients).
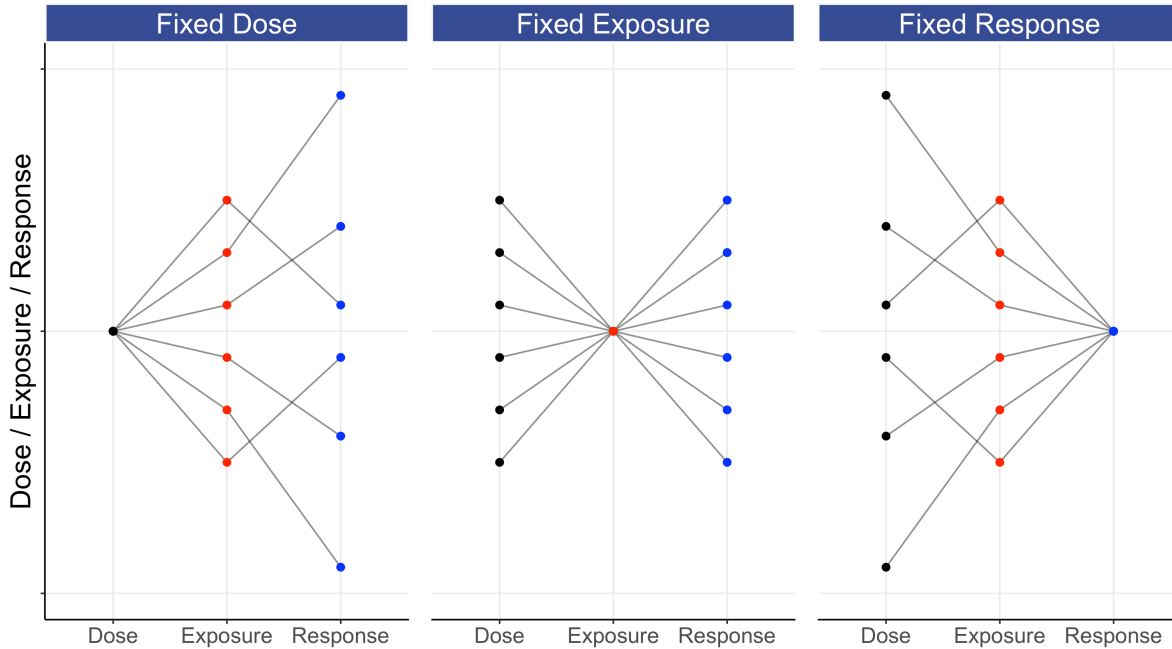
Figure 6.3: The relationships between dose, exposure and response under three strategies.

This figure may take a moment to fully grasp, since the y-axis is simultaneously quantifying Dose, Exposure and Response. In the left panel, all patients receive a "Fixed Dose", and this leads to a wide range of exposures (e.g. C_{ave}), that in turn lead to a very wide range of responses. In the middle panel, all patients now have the same exposure (for example, the dose has been titrated to achieve a given exposure). Here we see that a wide range of doses are needed to achieve the same exposure (due to IIV in PK), and that the same exposure leads to a wide range of responses (due to IIV in PD). In the right panel, we have the same responses for all patients (for example, the dose has been titrated to effect, either with or without the use of exposure measures). Note the cumulative effect of the variability as we travel from either dose to response (left panel), or from response to dose (right panel); stately alternatively, giving the same dose to all patients will lead to a wide range of individual responses, whilst obtaining the same response in all patients will require a wide range of individualized doses. Thus although giving the same dose to all patients is "simple", it is not what we seek; we seek to get the same (good) response in all patients.

Although clearly an oversimplification of the complex cascade from dose to response, the above illustrates a very important point:

> **!** Important
>
> **Dose is just a very crude mechanism by which we seek to deliver sufficient drug to the site of action to illicit the desired PD responses.**

Thus we should **always expect** to need to change the dose to achieve the best responses in all patients; we need **Personalised Dosing** because of the omnipresence of IIV in PK and PD.

It is useful to contrast how each panel in the figure above translates to a drug development strategy.

- Fixed Dose: What is the "optimal" one-size-fits-all dose.

- Fixed Exposure: What is the "optimal" target exposure (e.g. Therapeutic Drug Monitoring)

- Fixed Response: What is the "optimal" dose-titration algorithm.

As concrete examples, we can compare two drugs used to treat type 2 diabetes. The dipeptidyl peptidase-4 (DPP-4) inhibitor sitagliptin (brand name Januvia) is given to all patients at a 100 mg dose ("Fixed Dose"). The long acting insulin, insulin glargine (brand name Lantus), is titrated to effect using self-measured blood glucose levels ("Fixed Response"). With lifetime sales of >$30 billion dollars (sitagliptin) and >$60 billion dollars (insulin glargine), both drugs have been commercially very successful. Based on the pharmacology and mechanism of action (MoA) of each drug, both drugs were sensibly developed and commercialised. These drugs will be discussed furthermore in this book, but briefly 100 mg of sitagliptin is a very high dose, since this dose is approximately 20 times the ED50 (the dose which elicits 50% of the maximum drug effect) [3]. Importantly, DPP-4 inhibitors are remarkably "clean" drugs with excellent safety/tolerability profiles, allowing such a very high dose to be given to all patients. If patients fail to response adequately to 100 mg, it would be unwise to consider an even higher dose; rather, the patient just does not response well to drugs with this MoA. In contrast, too high a dose of insulin glargine can lead to hypoglycemia (very low blood glucose) and hence it is sensible that the patient slowly up titrates from an initial low dose. Insulin glargine is delivered subcutaneously using injection pens where the patient turns a dial on the pen to select the right dose (a brilliantly simple and effective dosing device). The ability for patients to tailor their insulin glargine dose to achieve the desired response is the key to the success of insulin glargine; it has been very successful **because** the dose can be titrated appropriately, and not **despite** the fact that the dose needs to be titrated.

Drugs like sitagliptin are very rare, and indeed we will refer to types of drugs herein as **diamonds**; they exist, but are scarce among our much more common **coals**. Most drugs are **coals**; they do require us to carefully balance the benefits and harms at the patient level.

7 Introduction To IIV In PK And Its Consequences To D-E-R Trial Design

At the end of this chapter, the reader will understand:

- The key parameters of a basic PK model.
- What "low", "typical" and "high" IIV in PK look like.
- The meaning of the terms linear and non-linear PK.
- How higher IIV in PK equates to wider distributions of average concentrations (C_{ave}) across individuals for a given dose, and hence has consequences for the optimal dose selection for population D-E-R modelling.

In the previous chapter we gave an initial introduction to IIV in PK and PD, showing a basic PK profile over time with repeat dosing, and a simple illustrative figure that demonstrated how the cumulative effect of IIV in PK and PD leads to wide ranges of responses across individuals.

We will now introduce IIV in PK in a way that is most relevant to drug development and clinical trial designs. The purpose here is **not** to explain all PK concepts, but rather to understand how the IIV in one PK parameter, clearance (CL), generally drives the IIV in average concentrations (C_{ave}) we see across individuals. Pharmacokineticists report the IIV in CL for drugs, so by understanding what is a "low", "typical" and "high" IIV in CL across drugs, we can better understand the general range on IIV in C_{ave} we expect in drug development, and the consequences for exposure response modelling and optimal clinical trial design. Importantly, it will show us how to "think" about dose range selection in our trials. That is, how the doses we select generate exposure ranges, and how these exposure ranges are critical when our goal is to accurately and precisely quantify D-E-R relationships.

For illustration, we will use one of the most basic PK models, the one compartment model with first order absorption model. For readers more interested in understanding IIV in PK from

a more general perspective, feel free to jump directly to the next figure, since the following formulae are provided to simply justify the foundations for what is shown in the figures.

$$Concentration = \frac{F \bullet Dose}{V} \left(\frac{ka}{ka - ke}\right) (e^{-ke \bullet t} - e^{-ka \bullet t})$$

Here F is the relative bioavailability, V is the volume of distribution, ka is the absorption rate constant, and ke is the elimination rate constant. You can read more about these pharmacokinetic terms on Wikipedia.

Clearance (CL) is defined as 'the volume of blood cleared of drug per unit time', and we have formulae that link CL to V, ke and half-life ($t_{1/2}$).

$$CL = V \bullet ke = V \bullet \frac{\ln(2)}{t_{1/2}} = \frac{0.693 \bullet V}{t_{1/2}}$$

The half-life, $t_{1/2}$, of a drug is the time it takes for the amount of a drug's active substance in the body to reduce by half, and the above shows that if there is an inverse relationship between CL and $t_{1/2}$; doubling CL halves the $t_{1/2}$. This points to how CL drives the average concentrations at steady state for this model, C_{ave_ss}, described in the equation below:

$$C_{ave\_ss} = \frac{F \bullet Dose}{CL \bullet \tau} = \frac{AUC_{ss(0-\tau)}}{\tau}$$

Where τ is the dosing interval (for example, every 7 days), and $\text{AUC}_{ss(0-)}$ is the "Area Under the Curve" over the dosing interval at steady state.

Finally, when F is constant across individuals, we see that:

$$ln(C_{ave\_ss}) \propto -ln(CL)$$

Hence IIV in ln(CL) is directly proportional to IIV in ln(C_{ave_ss}). Thus we can utilise the extensive knowledge pharmacokineticists have accumulated over the last 50 years around the IIV in clearance across drugs to broadly quantify the magnitude of the heterogeneity in average concentrations across individuals.

Figure 7.1 shows concentrations profiles over time for 16 simulated individuals following administration of a 10 mg dose each week for 6 weeks. The simulation uses our basic PK model with an IIV in clearance of 40% for this hypothetical drug (the total concentration is determined using the "superposition" principle whereby the total concentration is calculated as the sum of the concentrations from each dosing event (this implicitly assumes linear kinetics)).

The first observation is both the most simple and the most important; **although all individuals may receive the same dose, their concentration profiles over time will be**
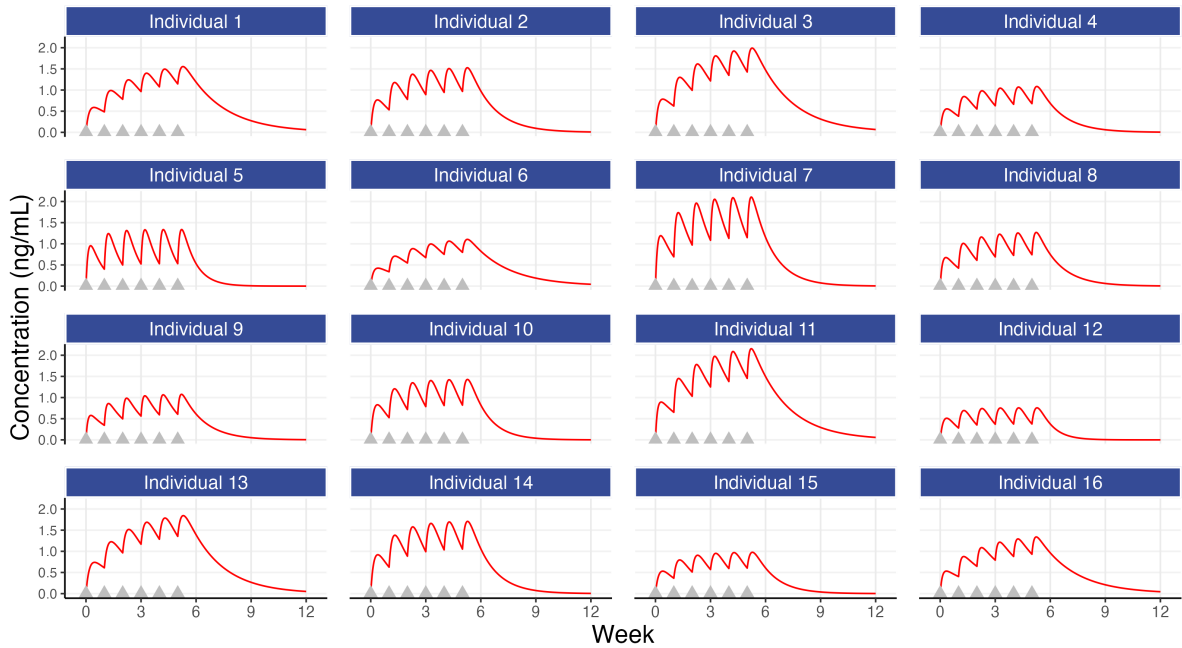
Figure 7.1: Concentration profiles over time for 16 simulated individuals following administration of 10 mg each week for 6 weeks. To aid the visual comparisons across individuals, all 16 individuals are shown in each panel in light red. The 6 doses are shown as gray triangles.

different (e.g. the concentrations for individual 7 are approximately 3 times higher than those for individual 12). This observation alone points to the expected necessity for different individuals to receive different doses, since it is generally these concentrations, and not the administered dose, that will be the primary driver of changes in the PD endpoints. The figure also shows that steady state is achieved in most individuals within 5-6 weeks, although again there are clear differences across individuals in how the drug concentrations change/accumulate over time with repeat dosing. In this simulation, as is real life, the individuals are heterogeneous, and we must always remember this when we think about the 'optimal' dose for each individual.

Figure 7.2 is similar to the above figure, but now showing the concentrations profiles for 100 individuals overlaid for three different magnitudes for the IIV in clearance (a "low" figure of 25%, a "typical" figure of 40% (as used above) and a "high" figure of 55%), on both the untransformed concentration scale (top) and log transformed concentration scale (bottom).
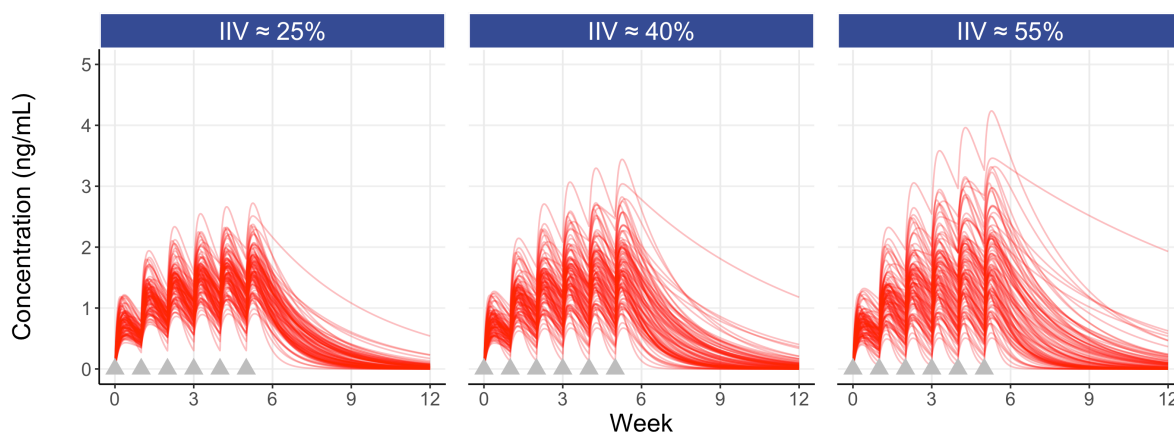
Here the influence of the magnitude of the IIV in clearance can be seen. A drug with a low IIV yields concentration profiles that are less variable across individuals, whereas a drug with a high IIV yields concentration profiles that are more variable across individuals (note: the IIV we observe is reflective of a given drug regimen (e.g. the drug, the dose regimen, the formulation, the route of administration etc.) and the patient population (comedications, type of disease etc.) which are typically "fixed"; that is, we generally **cannot** influence the magnitude of the IIV, but we **can** intelligently design our D-E-R trials using our knowledge of the expected IIV in PK for our drug).

Following the last dose at week 5, the figure also highlights (shaded gray area) the IIV in C_{ave} within the inter-dosing interval between weeks 5 and 6. Since dosing is then stopped at week 5 in this simulation, we see the elimination phase thereafter. Note that individuals with the **highest C_{ave}** typically eliminate (remove) the drug more slowly (these are the individuals with the **lowest clearances**).

Since this is a simulation, we can show the equality of the distribution of individual CL values (CL_i) used in the simulations to the distribution of the individual average concentration
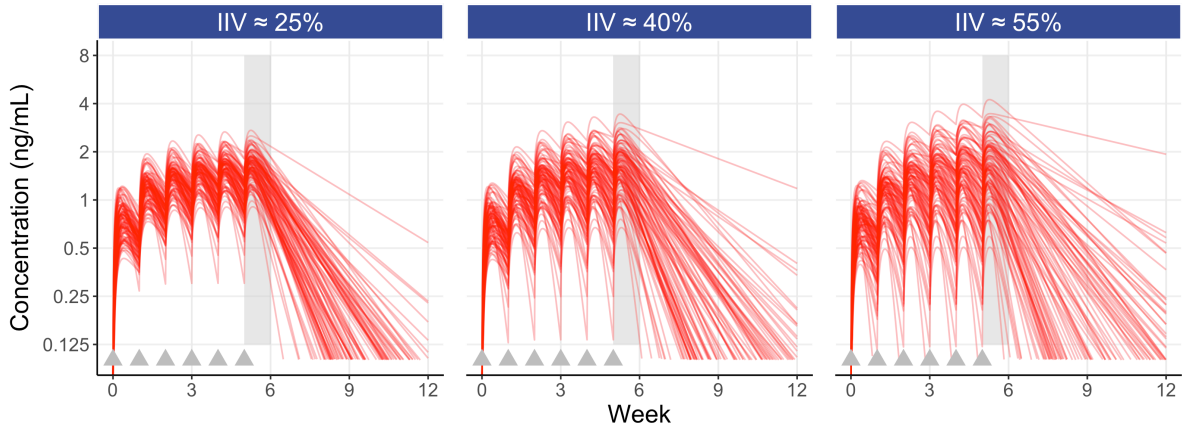
Figure 7.2: PK concentration profiles over time for 100 individuals on untransformed (top) and log transformed concentration scale (bottom) for 3 different levels of IIV in clearance.

between weeks 5 and 6 (C_{ave}) for the 3 levels of IIV for our 100 individuals. This is shown in Figure 7.3 below; note the (approximate) symmetry around the centre of the figure.

For the different levels of IIV in clearance, we can determine the ratio from the individuals with high exposures (i.e. the 97.5% percentile) to the individuals with low exposures (i.e. the 2.5% percentile). For an IIV of 25% this ratio is approximately 2.7 (exp(1.96*0.25)/exp(-1.96*0.25)), for IIV of 40% it is 4.8, and for an IIV of 55% it is 8.6, **so although individuals may be receiving the same dose, the heterogeneity in their average drug concentrations can be very large**.

When patients all receive the same dose but their actually drug concentrations vary enormously we should understand, from just a PK perspective alone, that the chance that the drug concentrations for each patient are exactly at the right level for them is infinitesimally small.

We have begun to understand how different levels of IIV in average concentrations can be motivated from a very well understood concept in PK, IIV in clearance. In our basic PK simulation, we have only considered so called **linear kinetics**, whereby the concentrations at steady state for an individual will be determined by their (constant) clearance and dose; if we double their dose, we double their concentrations. At some drug dose combinations, we may observe **non-linear kinetics**, whereby clearance is not constant within an individual. For example, a drug may be metabolised by a particular enzyme. For low doses/concentrations, the clearance will be constant, as there is sufficient enzyme to metabolise the drug. However as the dose/concentrations increase, the capacity of the enzyme to remove the drug will be rate limiting, thus lowering the clearance. In these more complex situations, we would generally expect **additional** sources of IIV beyond that observed with the simpler linear kinetics example considered above. Thus for the moment we will further consider our basic simulation, but will
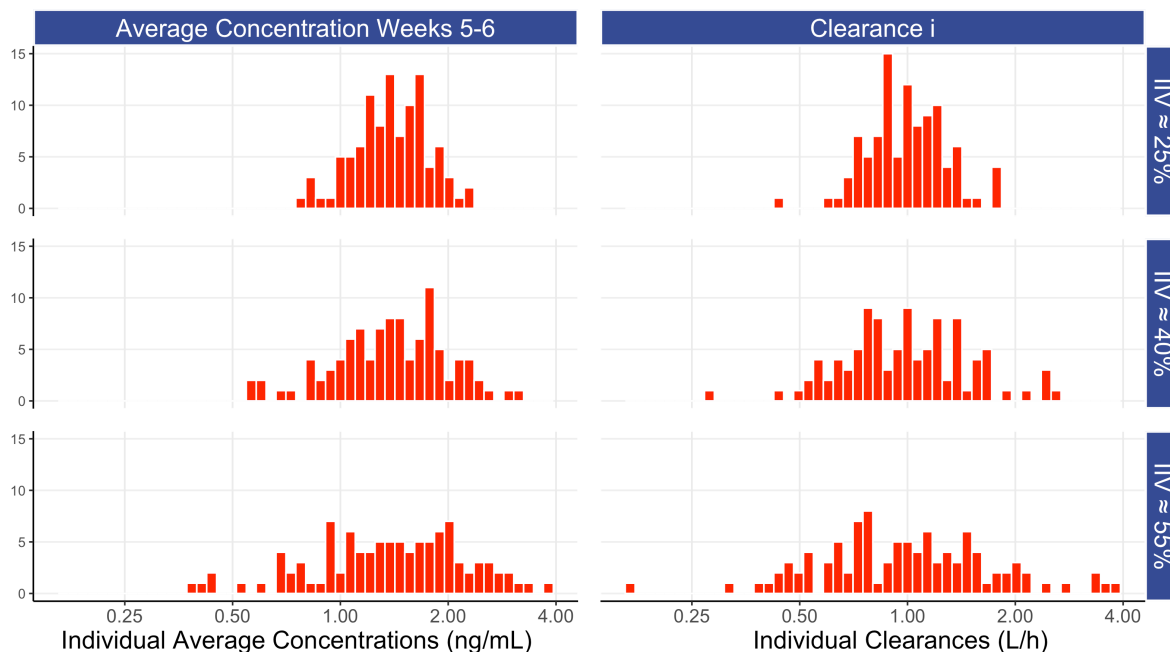
Figure 7.3: Distributions of average concentrations (weeks 5-6) and individual clearances across 100 individuals for different magnitudes of IIV in clearance.

remain mindful that each drug has its own specific ADME properties that will ultimately drive the IIV in PK observed at each dose across the dose range (pharmacokineticists are well skilled at quickly determining the IIV in PK from analyzing the early SAD and MAD phase 1 trials that employ frequent PK sampling).

We will now discuss the importance of IIV in PK with regards to how we interpret PD effects, and its centrality when seeking to determine the best design (i.e. dose levels, dose spacing etc.) for a particular type of D-E-R analysis (best design = most informative = highest precision on D-E-R relationships).

For our basic PK simulation, we will compare the different distributions of exposures (as measured by individual average concentration between weeks 5 and 6 (C_{ave})) across two trial designs where the designs are the same except for the dose levels. Each design will have n=100 individuals per dose level, with the following doses:

- Design 1: doses 5 mg, 10 mg, and 20 mg

- Design 2: doses 2.5 mg, 10 mg, and 40 mg

In Figure 7.4 below we show the distributions of 100 individual exposures (using C_{ave}) for each dose level, and combined across all dose levels for each of the two designs (for the "typical" IIV in clearance = 40 %).

Figure 7.4: Distributions of average concentrations across each dose level, and combined across doses for each design.

Understanding this figure is crucial to understanding many important concepts around the role of dose and how it leads to a range of exposures and ultimate responses.

> **!** Important
>
> **To most accurately and precisely quantify D-E-R relationships, we must "intelligently" select dose levels/ranges using knowledge of the PK of the drug.**

Please study this figure carefully and call me if it is still unclear!

Figure 7.4 provides a first insight in the potential relationships in **PD effects** we may expect between different doses. For example, for this "typical" drug we see a significant overlap in exposures between adjacent dose such as 5 mg and 10 mg and 10 mg and 20 mg. That is, it makes no sense to think about 5 mg and 10 mg as separate entities; they share so much of the same DNA so to speak. Equally, it would be bizarre to expect the 10 mg effect to not be somewhere between the effects seen at 5 mg and 20 mg. Thus whenever we see two closely spaced doses that yield overlapping exposure distributions, we can usually be confident the responses at the two doses will be reasonably similar. In later chapters we will discuss the steepness/shallowness of D-E-R relationships for **PD effects** seen in drug development, and hence will be able to understand/quantify how the degree to which the exposure distributions do or do not overlap between doses can equate to true differences in response rates between dose levels.

The figure also illustrates **how we need to select/evaluate dose levels/ranges when we design certain dose-ranging trials** (those where we wish to subsequently conduct a Population D-E-R analysis). In these analyses the goal is to combine data across all individuals and doses, but rather than link dose to response directly (i.e. a D-R model), we utilise the "intermediate" individual exposure measures and link these to the observed response (the E-R model). When we combine the data across doses for the planned E-R analysis, what matters is the **combined distribution of exposures across all doses achieved with the proposed design**. In the above figure, this is shown in the "**All Doses**" panel at the bottom for each design.

When we compare design 1 with design 2, design 1 has doses that are far too closely spaced, with the combined exposure distribution ("**All Doses**") having a lot of data near the middle of the exposure range (around 0.5-4 ng/mL), but much less data beyond this range. In contrast, design 2 has an exposure range that is approximately 4-fold wider compared with design 1 (since a 16-fold dose range (40 mg / 2.5 mg) was used, rather than the 4-fold dose range (20 mg / 5 mg)). In the next chapter we will see how the doses used in design 2 yield a much greater understanding on the true D-E-R relationship relative to the weaker design 1.

Before finishing the discussion on design 1 versus design 2 with respect to PK, it is important to note how the IIV in average concentrations for a particular drug will determine how narrowly or widely we should space the dose levels for subsequent D-E-R analyses. In the above example,

we chose a "typical" drug IIV of 40%, and briefly compared and contrasted two different designs. In this case, spacing the dose levels as in design 2 looks superior to design 1, as it yielded a wider exposure range whilst ensuring there were no "gaps" in exposure ranges from having too widely spaced doses (e.g. if doses like 1 mg, 10 mg and 100 mg had been used). If the drug has a **low IIV** in average concentrations, then generally the **doses need to more closely spaced** since the individual distributions of exposures from each dose will be narrower. In the case of drugs with **high IIV** in average concentrations, the opposite it true; the **doses can be more widely spaced**, since the exposure distributions from each dose will be wider (this is perhaps the only good thing about having a highly variable drug!). Clearly here we are only discussing the two designs in terms on the differences they yield in the combined exposures distributions (the "**All Doses**" panel), but what will be most important is the **location** of the resulting exposure distributions on the (PD) E-R curve. For example, in the above case if the "middle" of the E-R was at a C_{ave} of 100 ng/mL, both designs would be awful, as all doses in both designs are far too low. Thus when we formally introduce optimal clinical trial designs for D-E-R modelling, there will be a need to incorporate the expected E-R relationship, and hence allow the evaluation of different designs to "recover" the true E-R relationship as accurately and precisely as possible.

We will now turn our attention to IIV in PD, and again compare design 1 and design 2.

---

8 Introduction To IIV In PD And D-E-R Analysis As Evidence For Regulators

At the end of this chapter, the reader will understand:

- IIV in PK and IIV in PD **both** play a crucial role in the actual PD outcomes we observe across our (heterogeneous) patients.

- How each individual will always follow their own D-E-R curve; titrating their dose will move them along **their** curve.

- How E-R data/analysis is generally superior to D-R data/analysis.

- How better trial designs more precisely quantify D-E-R relationships. Stated equally, intelligently selecting the right dose levels to study reduces total sample sizes (N).

- If they truly wish to advocate dose optimisation to best serve and protect patients, regulators must state that well conducted trials that precisely quantify D-E-R relationships for efficacy and safety are **clearly superior** to the current practice of just looking to obtain two trials with P<0.05 for a single primary efficacy endpoint.

In the previous chapter we gave an initial introduction to IIV in PK. We will now give a similar introduction to IIV in PD.

Unlike PK, we have many different types of PD endpoints. These span the whole range of types of data, from continuous (e.g. blood pressure in hypertension), binary (e.g. fracture/no fracture in osteoarthritis), ordered categorical (e.g. headache recorded as none/mild/moderate/severe as an adverse event), count (e.g. number of seizures in epilepsy) and time to event (e.g. progression free survival in oncology). As such, it may not always be straightforward to think of each individual as having their own D-E-R relationships, but they do (even if we cannot always investigate them). Additional factors, such as the delay in observing the full PD effect for a given dosing regimen, may also serve to obscure these individual D-E-R relationships.

Herein we will extend the PK simulation from the previous chapter to drive changes in a PD endpoint. In later chapters the full details of this type of PD simulation will be introduced (i.e. the "link" between the PK and PD), but the salient points to understand at this early stage are:

1. The PD effect for each individual is driven by their individual PK concentration profile.

2. Each individual has their own D-E-R relationship.

3. There is no delay between the PK concentrations and the PD effect (a "direct" effect).

Although this is a simple example, the purpose here is to illustrate that we can understand individual PD effects in the same way as we understand individual PK concentration profiles. That is, the characteristics of each individual will result in their own D-E-R curves, coming from the IIV in PK and PD.

Figure 8.1 shows the PD response profiles over time for 16 simulated individuals following administration of a 10 mg dose each week for 6 weeks. The simulation uses the individual PK concentrations to "drive" the individual PD responses.

Note: for technical experts, these individual D-E-R curves are based on the sigmoidal Emax model with IIV (random effects) on Emax (mean 100, SD=20), ED50 (mean 1 ng/mL, CV=50%) and Hill coefficient (mean = 1, CV = 20%).

Like in the PK example, the individual PD profiles are variable across individuals. The source of this variation is now composed of two parts:

- The IIV in PK leading to different concentration profiles over time across individuals (the D-E part)

- The IIV in PD leading to different responses to concentration across individuals (the E-R part)

The first observation is again the most simple and the most important; **although all individuals may receive the same dose, their PD profiles over time will be different (e.g. the PD responses for individual 5 are much higher than those for individual 15).** Thus for some individuals a 10 mg dose may be far too high or far too low. For example, if we assume that we would want a response of 50 or more by week 6 for each of the individuals shown above, it would be clear by the end of week 1 that individual 15 is not on the right path with this dose (so should we "plough on regardless" with 10 mg, or increase the dose if tolerability/safety data suggests it is reasonable to do so?).

Figure 8.2 shows the individual PKPD relationships (the E-R) for the 16 simulated individuals.

Figure 8.2 shows that each individual has their own E-R curve. For example, individual 5 is more sensitive to the drug, whilst individual 15 is less sensitive to the drug. The dose will determine which part of the E-R curve is covered for each individual (e.g. individuals 3, 7 and
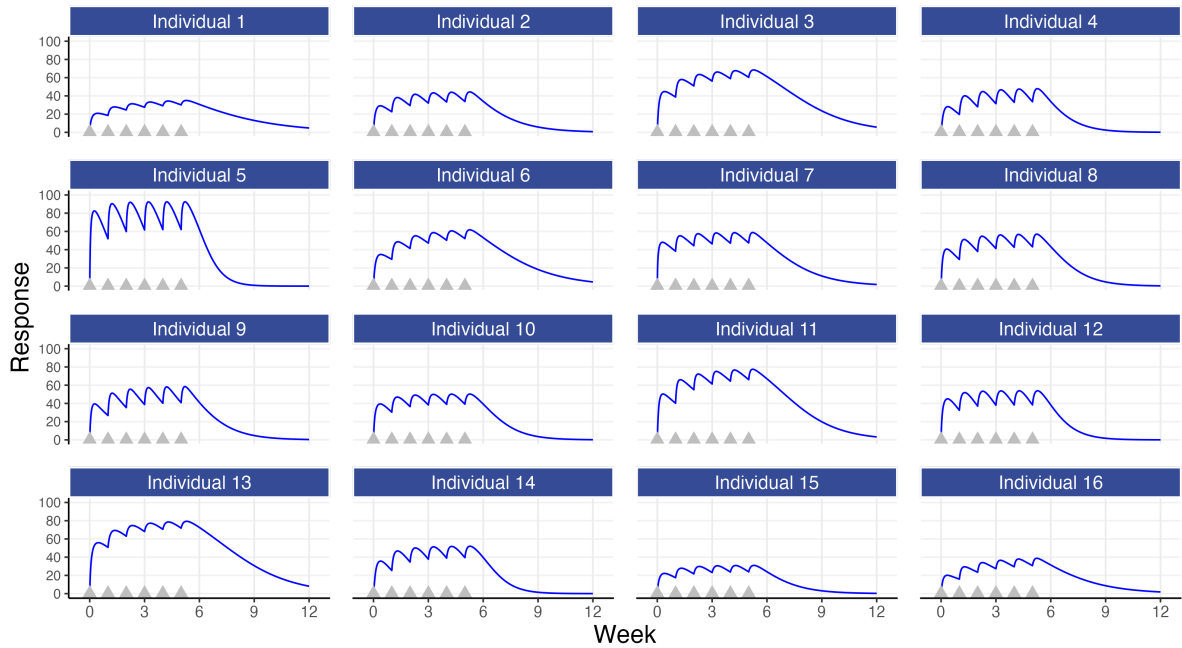
Figure 8.1: PD profiles over time for 16 simulated individuals following administration of 10 mg each week for 6 weeks. To aid the visual comparisons across individuals, all 16 individuals are shown in each panel in light blue. The 6 doses are shown as gray triangles.
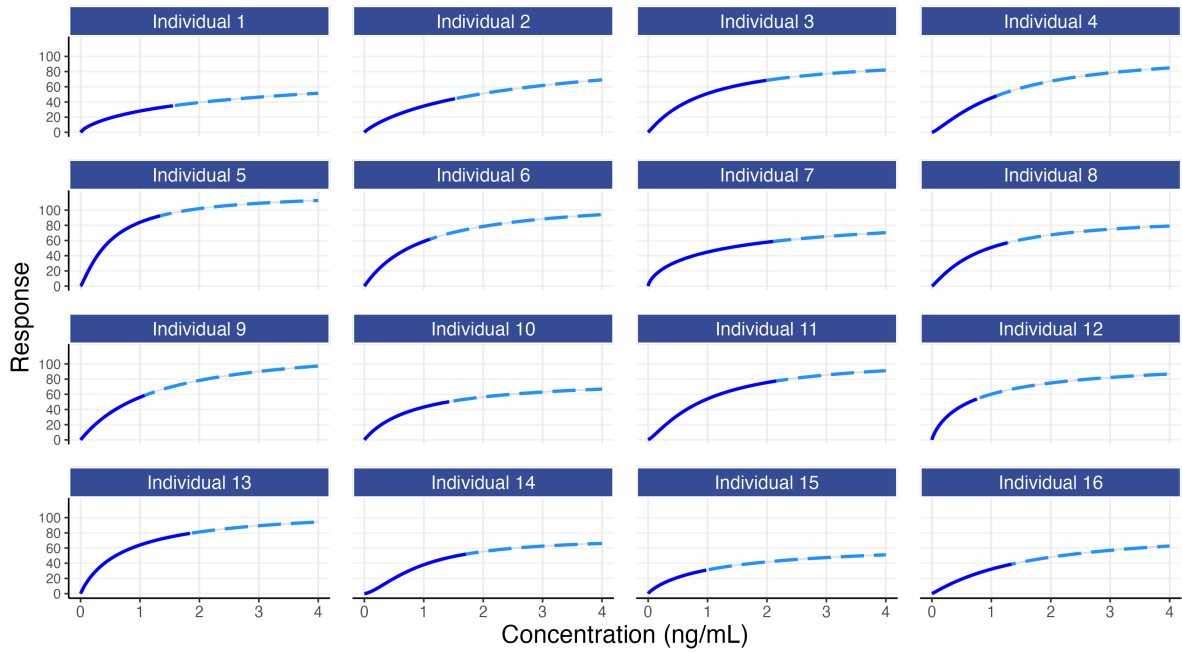
Figure 8.2: PKPD relationship for the 16 simulated individuals. The solid blue line shows the E-R range achieved with the 10 mg dose. The dashed blue line shows the true E-R relationship for each individual (as a visual reference, the gray lines show the E-R for all 16 individuals extended up to 4 ng/mL).

11 achieve concentrations up to around 2 ng/mL with the 10 mg dose, whilst individuals 12 and 15 only achieve concentrations up to 1 ng/mL with this dose).

The corresponding PK profiles for the 16 simulated individuals (as shown in the previous chapter) are shown below. You can switch between the 3 tabs at the top to move between the PK, PD and PKPD plots for these 16 simulated individuals.

8.1 PK
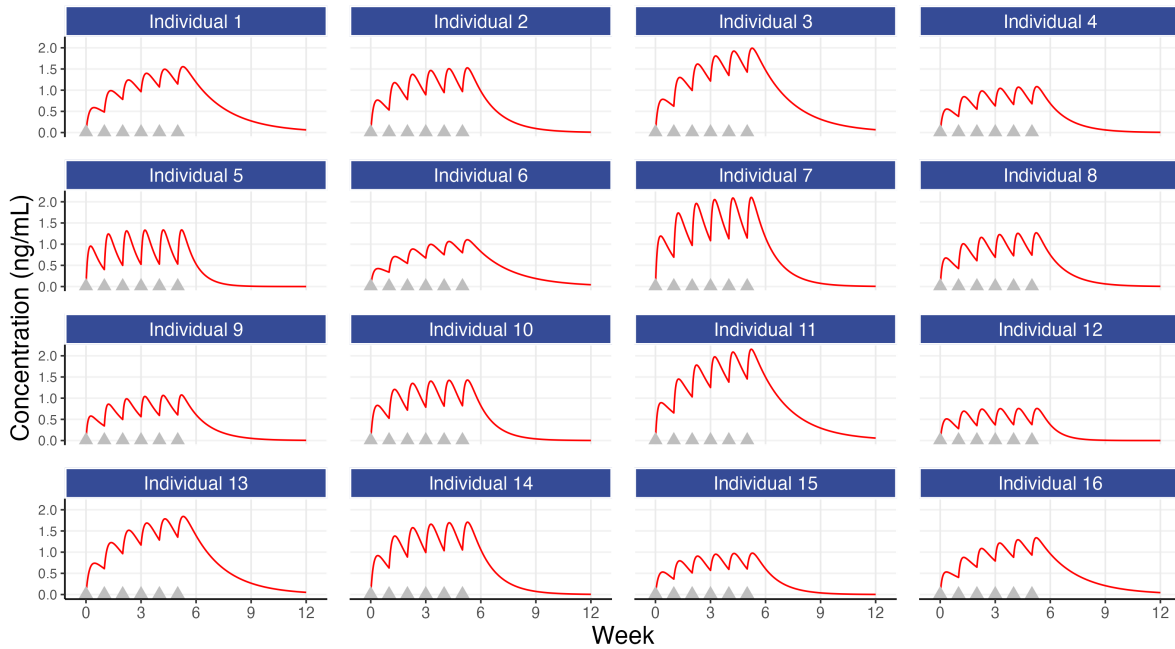
8.2 PD

8.3 PKPD
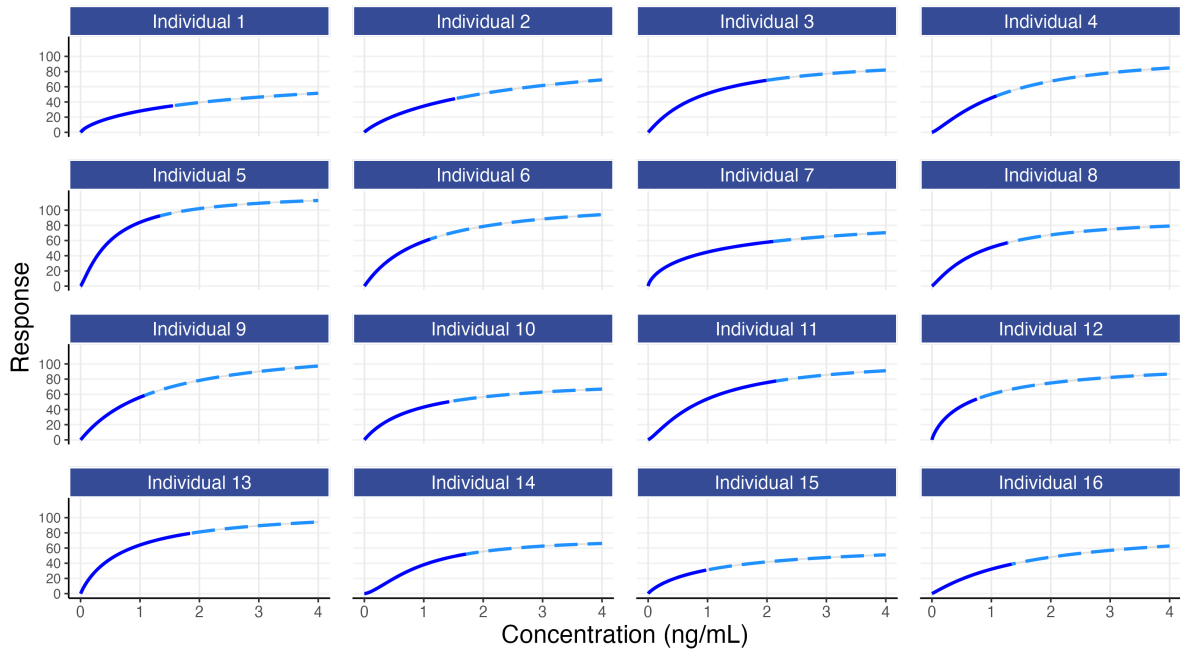
We can examine the figures above and make some observations. We see that individual 15 had generally low concentrations that, combined with their modest concentration response relationship, yielded a poor PD response relative to the other individuals. There are also individuals with very good outcomes since they had both higher concentrations **and** were more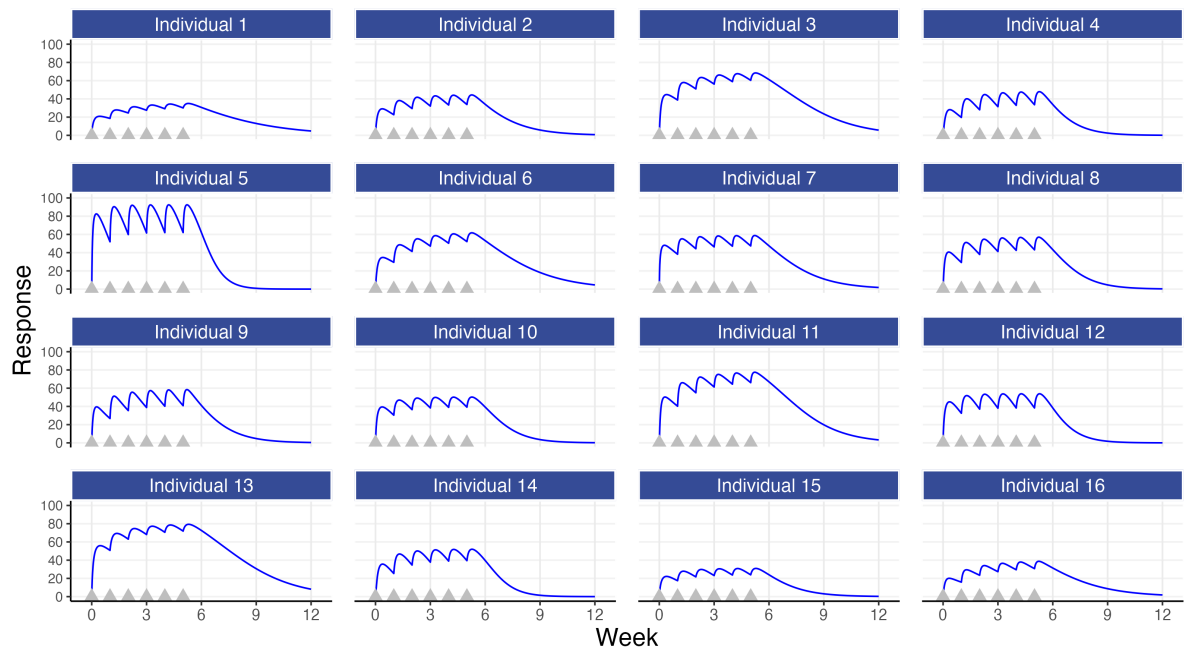 sensitive to the drug (e.g. individuals 3, 11 and 13) but also individuals with 'average' concentrations but still very good outcomes (e.g. individual 5 and 6) and individuals with 'average' concentrations but with poor outcomes (e.g. individuals 1 and 16).

❗ Important

IIV in PK and IIV in PD **both** play a crucial role in the actual PD outcomes we observe across our (heterogeneous) patients.

We may ask, how would individual 15 have fared with doses other than 10 mg? This is shown below in Figure 8.3 for fixed weekly doses of 10 mg or 20 mg or 40 mg or 80 mg each week, and with a weekly titration from 10 mg to 20 mg to 40 mg to 80 mg for the first 4 doses (e.g. assuming the responses at weeks 1-3 where not considered sufficient **and** the tolerability/safety were considered acceptable to consider higher doses).

Figure 8.3: Concentration (top) and response (bottom) profiles for individual 15 with doses of 10 mg, 20 mg, 40 mg and 80 mg and (last panel) following a simple weekly dose titration. The 6 doses are shown as gray triangles.
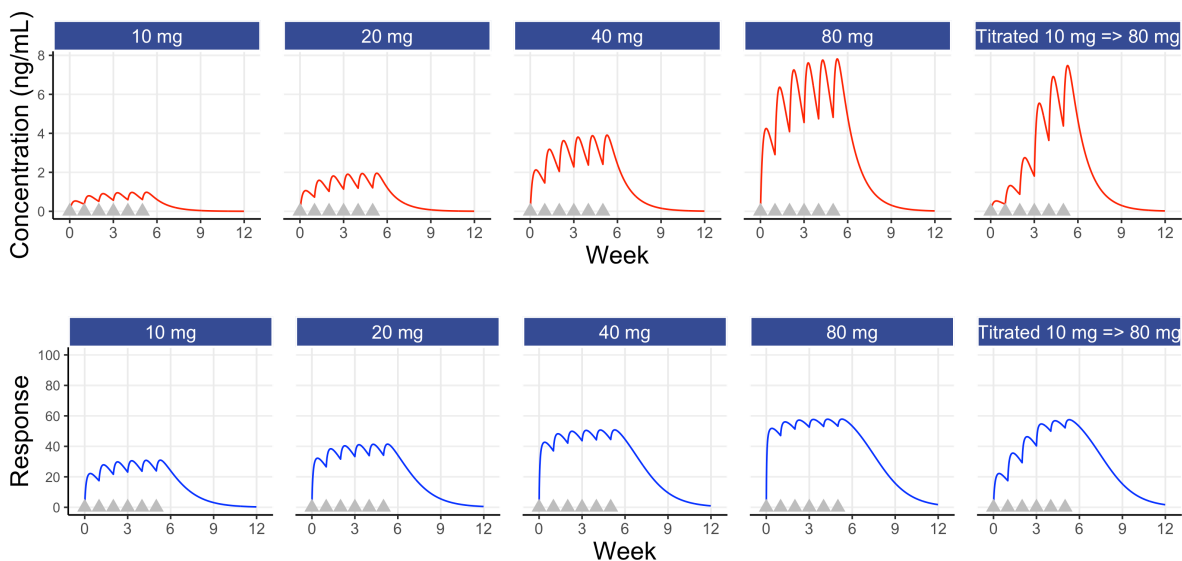
Here we see that individual 15 would need a dose of 40 mg or 80 mg to achieve a response of 50 or above at week 6; 10 mg is not the right dose for this individual! Since, a priori, we do not normally have the patient characteristics than will tell us **which patients** will need **which doses**, it seems clear that even a very basic titration algorithm, like that shown above, could be used to achieve **Personalised Dosing** (note: when the PD effects are delayed relative to the dose changes, we just need to wait sufficient time before making dose changes). For example, the dose is titrated upwards when/if the patient considers the previous dose as tolerable to them and they would wish to try the higher dose.

This basic simulation can be extended to discuss the same trial designs as considered in the previous chapter, but now with a focus on the PD responses. Recall our two simple designs, each with 100 individuals per dose.

- Design 1: doses 5 mg, 10 mg, and 20 mg

- Design 2: doses 2.5 mg, 10 mg, and 40 mg

For each individual we can now calculate a PD measure, such as the average response over weeks 5-6, and plot this against dose or against their average concentration at week 5-6 (C_{ave}). These are shown below in Figure 8.4. By summarising the data in this way (i.e. one PD measure and one PK (exposure) measure per individual for a single dose), we can more generalise our discussions to the common case where, for each individual, a single PD response is measured at some key time point (e.g. week 12 or week 26) and linked to either the dose or an appropriate single measure of exposure (e.g. C_{ave}).

54

Figure 8.4: Average response (weeks 5-6) versus dose (top) or average concentration (bottom) for each design. Colour changes by dose. Each dot is one individual. When plotted versus dose, individual data is 'jittered' to better visualise the data, and a standard boxplot of the data is overlaid.

The above figures are called the **population D-R** and **population E-R** respectively. There are many important observations that should be noted from the above two figures.

- For the D-R data (top figure), we see the PD responses across the 3 dose levels with design 1 are not well separated and any D-R analysis of this data would be highly questionable/imprecise; this is not a good design as the doses are too closely spaced. Design 2 has a greater separation in responses between the 3 dose levels, but understanding what is happening between the doses (e.g. 5 mg) will be dependent on the model used to link the doses, since we are "blind" to the true shape of the D-R between the 3 dose levels studied.

- The E-R data (bottom figure) is much more informative than the D-R data, since some of the variability we see in the PD variability is now being explained by differences in

the PK (i.e. we can see an E-R relationship both across doses **and within each dose**, whereas with the D-R all of these PD responses are (naively) collapsed to a single point on the dose scale). Again design 2 is superior to design 1, since the wider exposure range will result in greater precision when the E-R data in analysed. However even with the 16 fold dose range with design 2, it is questionable whether we would confidently conclude that the minimum and maximum effects are, as used in the simulation, 0 and 100 respectively. However the data generated from design 2 is clearly superior than that for design 1 (i.e. better designs are more informative for the same N).

- The distribution of concentrations for each dose with the E-R data is driven here by the IIV in clearance (here we use our "typical" value of 40%). A drug with a **higher IIV** in clearance would have **more** overlap between the doses, and a drug with a **lower IIV** in clearance would have **less** overlap between the doses. Thus we must use our PK knowledge on the expected distribution of exposures at each dose to (optimally) select the most informative dose levels and ensure no "gaps" in the exposure range; **this will allow excellent predictions across the whole dose range, in this case from 2.5 mg to 40 mg.**

- This simulation has **no measurement error** in either the PK or PD measurements. Every point we see in Figure 8.4 is a single point on that individual's true D-R or true E-R curve (recall each individual has their own D-R and their own E-R (as shown in Figure 8.2)). The variability we see in the PD responses (y-axis) reflects the heterogeneity in the **true** individual responses for a given dose. In the real world, when we review such figures, we must fully appreciate that each point is one observation (true value + random measurement error) from each **individuals** D-R or E-R relationship and not, as some may incorrectly assume, only random measurement error away from some true D-R or E-R relationship that is identical for all individuals. **Individuals will always follow their own D-R and E-R curve, and not the population D-R or E-R curve.** To emphasis this important point, the above figures are repeated below in Figure 8.5, but now with the individual D-R or E-R relationships shown from 2.5 mg to 40 mg. Thus population D-R and E-R relationships like those shown above (determined from fixed-dose parallel group dose-ranging trials) are of limited value when we discuss how best to titrate a dose **within** an individual.
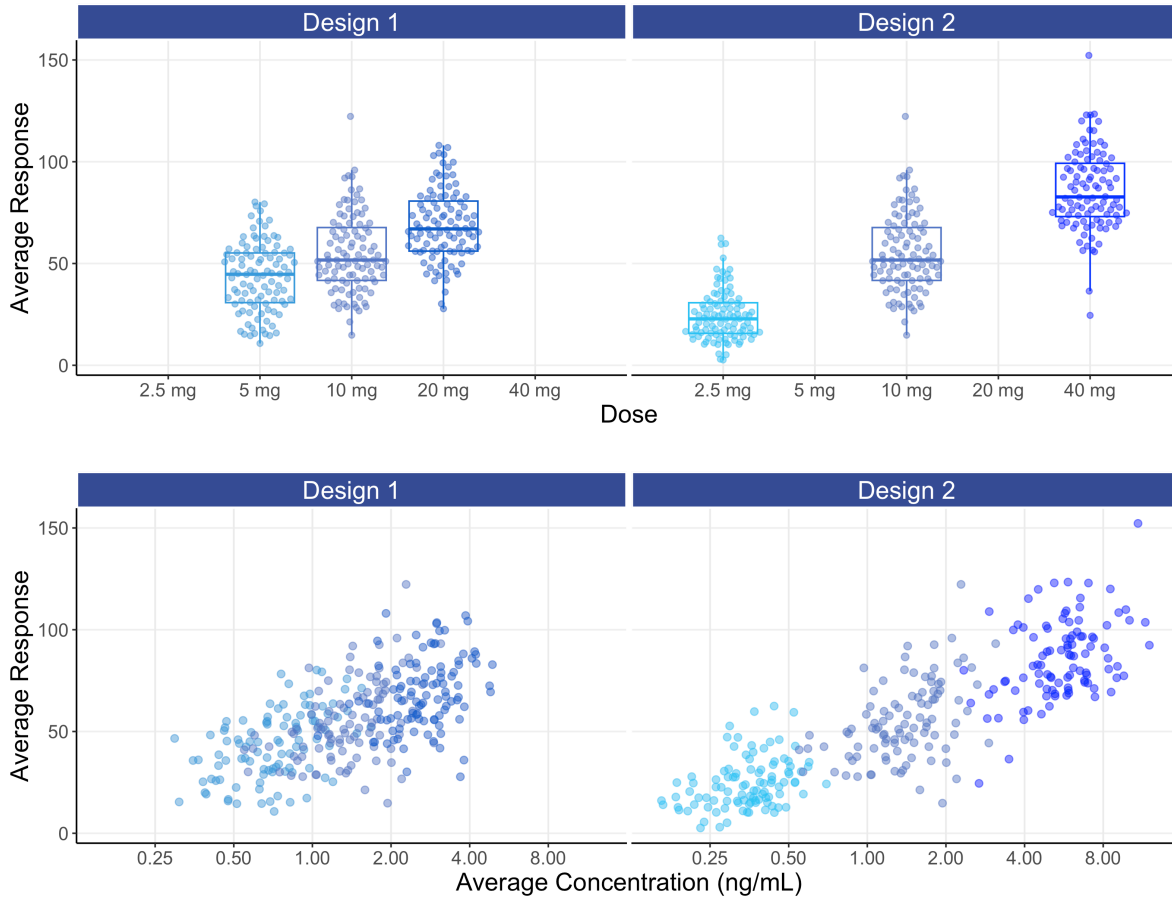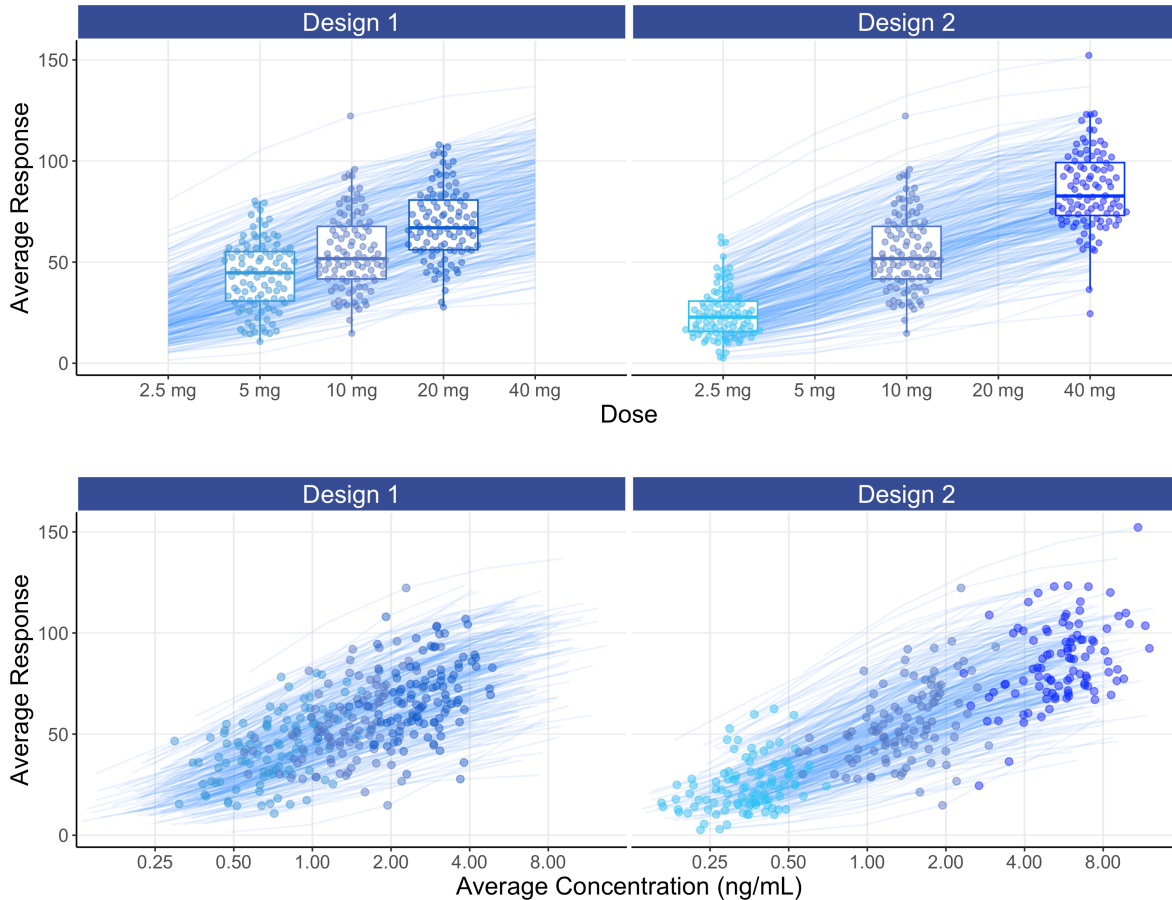
Figure 8.5: Average response (weeks 5-6) versus dose (top) or average concentration (bottom) for each design. Colour changes by dose. Each dot is one individual. When plotted versus dose, individual data is 'jittered' to better visualise the data, and a standard boxplot of the data is overlaid. Lines show the individual D-R or E-R relationships.

In general, it will be better to favour an E-R analysis ahead of the simpler D-R, as we are taking advantage of the additional information for each individual that the PK measure provides. Here we see the E-R data providing much greater granularity and insight beyond using just the 3 points (doses) in the D-R model (note: the E-R model is easily combined with the PK model to construct the full D-E-R relationship). With only 3 dose levels, understanding the true shape of the D-R relationship is very difficult, as the model may go perfectly through the average response for each of the 3 points, but there is no way to assess any "lack of fit" (that is, the extent to which the fitted model does/doesn't describe the data). In contrast with the E-R model we have, for example in this case, 300 unique exposures values (100 individuals * 3 doses levels) on the x-axis, and can therefore better assess the quality of the fitted E-R model across the whole exposure range.

In this example, the wider exposure range with design 2 will allow the analyst/regulator/company to more accurately and precisely characterise the full E-R, and hence D-E-R, relationship compared with design 1.

Note how design 2 more successfully achieves a continuum of exposures across the whole exposure range relative to design 1. This illustrates one of the clear benefits of such an E-R analysis, that of being able to confidently predict all doses within the dose range investigated, and not only the actual doses considered. That is, under design 2, the predictions at 5 mg (a dose not investigated) would be perfectly as sound as those at 2.5 mg and 10 mg (doses that were investigated), since design 2 clearly generates exposures that fully cover the exposure range that would be obtained at 5 mg.

> **!** Important
>
> An E-R analysis (rightly) forces a (smooth) continuum across the response data across all doses, which makes scientific sense.

In later, more technical, chapters we will cover how to optimally choose dose levels to best characterise this type of D-E-R relationship, but broadly speaking we need data at each part of the E-R curve: near the bottom (i.e. like placebo), either side of the middle of the E-R (where the E-R is steepest), and at the top of the E-R curve.

Thus when one understands how the data from all doses from well designed trials can be combined within an integrated E-R model to provide a comprehensive and coherent analysis, the idea of "cherry picking" a particular dose because its **observed** benefit/risk data **happened** to look a little more positive that adjacent doses looks both improper and unscientific.

> **!** Important
>
> Integrated D-E-R analyses, using all data from all doses, generate a **clearly superior** evidential basis for determining accurately and precisely how the measures of efficacy and safety/tolerability **truly** change as a function of the dosing regimen when contrasted with the simplistic "by trial, by dose" tables and listings.

Being fixated on **observed** outcomes for individual doses in individual trials is misguided. This is particularly problematic for weak designs that purport to be "dose optimisation" trials but only consider, for example, a "low" and "high" dose regimen. **These are never acceptable designs for dose optimisation; their use suggests a failure to prospectively assess the accuracy and precision for the planned D-E-R relationships.** This is not just my opinion. It is relatively simple to prospectively simulate (say 1000 times) such trials using **reasonable assumptions** about the expected D-R or E-R relationship, and then see how useful/useless such a trial would be. These simulations may show a "random" set of outcomes, with some simulations showing that the "low" and "high" doses are similar, and some where the "high" dose looks much better than the "low" dose. **Such simulations**

are essential to avoid conducting weak, and hence unethical, trials. In addition, further simulations that consider more dose levels over wider dose ranges would be shown to be more informative and useful for subsequently determining the D-R and E-R relationships. When crude "high" dose and "low" dose trials are run, the authors will often be incapable of meaningfully estimating any D-R or E-R relationships, and will therefore resort to simply "eye-ball" the observed data for each dose as though it is measured without error; they are only kidding themselves that such a weak and unscientific approach is truly "dose optimisation".

In contrast, **when integrated D-E-R designs/analyses are fully understood and embraced, they can provide much greater insight and flexibility to decide suitable dose ranges for approval based on the totality of the final evidence generated**. There is no need to "guess" acceptably "safe" dose(s) based on weak small phase 2 trials, but rather await sufficient data (N) across a very wide dose range to truly understand the full picture. Only then can both the pharmaceutical company and regulators determine the most suitable dose range for approval which may, or may not, coincide with actual dose levels used in the supportive trials.

The role and responsibilities of our regulators will be discussed in later chapters, but if they truly wish to advocate dose optimisation to best serve and protect patients, they need to stress the immense value and insight that D-E-R trials and analyses provide.

> **❗ Important**
>
> It is very important that regulators state that well conducted trials that precisely quantify D-E-R relationships for efficacy and safety are **clearly superior** to the current practice of just looking to obtain two trials with P<0.05 for a single primary efficacy endpoint.

As an final comment, I worked as an external consultant for a company and performed integrated D-E-R analyses for key efficacy endpoints exactly like that described above (integrated = all trials, all doses, all data). Throughout the development of the drug from phase 1 to phase 3, the company exclusively used these integrated D-E-R analyses (for both efficacy and safety) for **all** internal decision-making and dose selections. However at submission time, they presented the simple "by trial, by dose" tables, listings, and P values to the FDA advisory board; I presume because it was considered a "safer/simpler" strategy; this needs to change. Had the FDA specifically asked for the integrated D-E-R analyses, we would not only have seen the beautiful D-E-R analyses at the advisory board meeting, but importantly these results would have facilitated a more quantitative and coherent discussion around the benefits and harms across the whole dose range (potentially supporting a dose range to be approved). **Tables with P values do not achieve this!** I find in truly odd that companies can be doing advanced analyses to direct their own scientific decision-making, but seem reluctant to show these analyses to the regulators. We need regulators to demand such integrated D-E-R analyses are well planned, well conducted **and** made available for discussion at sponsor/regulator meetings. These critical analyses cannot be side-stepped or ignored. In addition, if a pharmaceutical company wishes to conclude that **"no D-E-R relationships could be determined"** for

key efficacy and tolerability/safety endpoints (due to their failure to study a sufficiently wide dose range and/or insufficient sample size), I think the company should consider designing and conducting additional trials, and **not** seeking approval when their dose regimen justification is wholly absent (good companies who well design their D-E-R trials will not end up in such an awkward situation!). Regulators need to be tough here (to ultimately protect patients).

---

9 Personalised Dosing; Patients Are Different

At the end of this chapter, the reader will understand:

- Why patients are a highly heterogeneous group of individuals.

- Understand adherence, and why it is important both for the patient and the pharmaceutical industry, but for difference reasons.

- Why we need very wide dose ranges if we wish to truly maximise patient outcomes (and hence minimise "churn").

- Why patients should be able to make decisions around their dose with consideration to their treatment goals, outcomes and personal preferences.

- That the consequences of getting the dose wrong goes well beyond the patient; their family, employer and society also suffer.

In previous sections we have introduced the idea that individual patient responses should be central to how we develop and use drugs, briefly covered the science of how doses lead to exposures that lead to responses, and that crucially patients have their own individual D-E-R relationships due to IIV in both PK and PD. Here we discuss this heterogeneity across patients in a much broader sense, and the consequences of this to adherence and the commercial success of the drug for the pharmaceutical company.

Patients are different. Not only do patients differ in simple measures such as age, weight and sex, each patient will have complex medical histories for the indication the drug will be used for. They may also have multiple comorbidities, and be taking a wide range of medications. In addition, they will have different lives, different values and different perspectives.

> **!** Important
>
> **Personalised Dosing** is about recognising this heterogeneity between patients and enabling each patient to find the right dose for them.

It is worthwhile considering many of the aspects around personalised dosing, and the consequences of getting the dose wrong.

Adherence may be defined as the extent to which patients take the drug regimen prescribed to them. If a patient considers their dosing regimen unacceptable in any way, they may decide to stop taking it. This may be due to many reasons, including the lack of efficacy, poor tolerability/safety, or finding the dosing regimen too burdensome. For both the patient and the pharmaceutical company, this failure is significant, but for very different reasons. For the patient, the drug regimen has failed to deliver, and they have wasted their time, money and perhaps experienced terrible adverse events. For the pharmaceutical company, they have lost a "customer" who will not return. They will not receive any future income from this patient, and the patient may tell other potential "customers" of their disappointing experience.

In a related conversation with a senior executive at a pharmaceutical company, we discussed the sales of one of their recently approved drugs. Two fixed doses were approved based on trials where patients typically received placebo, the lower dose (X mg) or the higher dose (2X mg). The indication, although not life threatening, would require the patient to continue taking the medication throughout their life (and there are very limited treatment options available in this indication). He mentioned that the company was disappointed that over 90% of patients prescribed to their drug had already stopped taking the drug at one year. From a tolerability perspective, I would describe the two approved dose levels qualitatively as "high" and "very high". The two dose levels only spanned a 2-fold range with overlapping exposure ranges, and hence the tolerability profiles for each dose were not well separated; both were poor. Given that there was no urgency to aggressively dose each patient immediately, and that finding the right dose for each patient could deliver a "lifetime" of sales, it seems obvious (at least to me) that having a very wide of doses (perhaps starting at X/5 mg) could have allowed each patient to tailor their dose using an informed dose-titration algorithm. Since the trials were never done, we can only speculate that such an approach would have yielded greater adherence, better patient outcomes and higher sales, but the significant consequences/failures of "**one-size-fits-all**" dosing must be recognised by commercial teams and senior management within the pharmaceutical industry. If a pharmaceutical company chooses to only offer a single dose level (or two very similar dose levels) there is a substantial risk that many patient will "slip between their fingers" because they failed to offer sufficient flexibility with respect to dose.

The term "churn" is used in business as a measure of the number of customers who leave in a given time period. Thus in this example, the "churn" equates to losing 90% of patients within one year. This is wholly unacceptable to both patients and the pharmaceutical company, and "discovering" this post approval is truly awful. From a purely economic perspective, the pharmaceutical industry should seek to minimise "churn". Fundamentally, any drug label with 1-2 dose levels will never enable personalised dosing. We need much wider dose ranges if we wish to truly maximise patient outcomes (and hence minimise "churn").

It is also valuable to discuss how patients will differ in their attitudes to their outcomes on a given dosing regimen.

Consider Lilly and Oscar, two epileptic patients both experiencing, on average, 4 seizures a week prior to treatment. After an initial period of dose titration, both Lilly and Oscar are on the same dose and now only experiencing 1 seizure per week, but with intermittent episodes of nausea and vomiting that they both report as "mild". Assuming they are in the middle of the approved dose range, should they continue on the same dose, reduce the dose, or increase the dose? This is a question we cannot answer, but the patient can. We should not pretend to know what is best for them, since this question directly relates to **their** experiences with the current dose, and **their** own attitudes to the perceived benefits and harms of the dose change. For Oscar, he may wish to lower the dose, as the "mild" nausea was actually quite debilitating for him, whereas Lilly might be keen to test the high dose, to see if further reductions in her seizure frequencies could be realised. This trade off between benefits and harms is often labelled the **utility**, and patients will differ in utility even when, on paper, they report identical outcomes. Thus Figure 6.1 can be augmented with an additional source of IIV, the variability between patients in their assessments on the utility of their responses. As drug developers, we should ensure a sound scientific framework for the approved dose range, but recognising patients interpretation of "best" for them will inherently be a personal decision (reached in dialogue with their physicians).

As a second example, consider personalised dosing for a drug approved to treat rheumatoid arthritis (RA). As a primary endpoint in regulatory trials, approval in RA is often based on the proportion of patients who achieve ACR20 (ACR50), a composite scale where the patient needs to have improved by 20% (50%) in 5 of 7 domains, with these domains spanning objective and subjective measures of pain and inflammation. What do the improvements in the ACR20 (ACR50) actually mean for individual patients? What dose is best for a particular patient? One RA patient may consider their knee pain most important since it stops them taking daily walks with their dog, whilst another RA patient may consider their hip pain most important since it stops them driving to see friends. Whether a patient can find a dose that works for them may be based on many factors that we do currently captured in our clinical trials, but simply asking the patient "Would you like to continue on this dose, or try a higher or lower dose?" would enable **the patient to consider their dose in consideration to their treatment goals, outcomes and personal preferences.**

The importance of personalised dosing extends beyond the patient and the commercial value of continued adherence to the pharmaceutical company. When patients are poorly treated, the societal impact cannot be overstated. In RA and epilepsy, and indeed most therapeutic areas, if patients are poorly dosed, they will continue to experience significant difficulties in "life", such as needing to take time off work, or needing to stop work all together. They may struggle to fulfill responsibilities at home and with their family, and their personal relationships may suffer. **That is, the consequences of getting the dose wrong goes well beyond the patient; their family, employer and society also suffer.** Similarly, when patients are over dosed, they may experience (severe) adverse events that could require hospitalisation, and lead to many of the same challenges with under dosing. Again, their family and society pay a price for this dosing failure. Equally, when we get the dose right for the patient, the benefits

extend well beyond the patient. Thus getting the dose right for each patient is absolutely paramount.

I hope the importance and value of Personalised Dosing is fully appreciated by all stakeholders: governments, regulators, industry, patient advocacy groups and patients.

---

10 Where Does Precision Medicine And Personalised Medicine Fit In?

At the end of this chapter, the reader will understand:

- What is meant by the terms precision medicine/personalised medicine.
- How precision medicine will help further refine inclusion/exclusion criteria for clinical trials.
- That precision medicine is both old and new; a new name for something that has existed for a very long time.
- That precision medicine will not be able to determine the right dose for each and every patient.
- Our patient level data is ever increasing both prior to, and during, drug treatment. Our goal must be to actually utilise this information to deliver better outcomes for patients.

When discussing **Drug Development For Patients**, we need to cover two terms that are frequently used in the discussion around individualized treatment regimens; precision medicine and/or personalised medicine. I see these terms used interchangeably, so for the remainder of this chapter I will only refer to precision medicine.

In 2015, Barack Obama famously launched the "Precision Medicine Initiative". Under the banner "It's health care tailored to you", it stated:

> *"Until now, most medical treatments have been designed for the 'average patient.' As a result of this 'one-size-fits-all' approach, treatments can be very successful for some patients but not for others. Precision Medicine, on the other hand, is an innovative approach that takes into account individual differences in people's genes, environments, and lifestyles. It gives medical professionals the resources they need to target the specific treatments of the illnesses we encounter, further develops our scientific and medical research, and keeps our families healthier."*

This is excellent. Clearly we should always strive to use individual patient characteristics to better tailor drug regimens. Similar statements about precision medicine have also been made in the mainstream medical literature [8], such as:

> *This identification and use of patient-level characteristics to deliver better outcomes for patients is often termed 'precision medicine' "*

The astute amongst you will recognise that this is nothing new. We have been using individual patient characteristics to tailor individual treatment regimens for a very long time. For example, for over 70 years we have been using a patient's blood type to ensure transfusions of compatible blood, or using a patient's body weight to determine the dose of a given drug [9]. Indeed the role of pharmacogenetics is not new. In 2004, Lesko and Woodcock [10] wrote:

> *"Pharmacogenomics and pharmacogenetics provide methodologies that can lead to DNA-based tests to improve drug selection, identify optimal dosing, maximize drug efficacy or minimize the risk of toxicity."*

Precision medicine has also been defined as a tool for selecting patients for clinical trials who may be expected to benefit most. For example, Cook [11] wrote:

> *"Precision medicine is an approach to developing drugs that focuses on employing biomarkers to stratify patients in clinical trials with the goal of improving efficacy and/or safety outcomes, ultimately increasing the odds of clinical success and drug approval."*

Again, this is nothing new. We have frequently selected specific patient sub-populations based on the expectation that they will benefit most for the given drug regimen. For example, in the development of statin therapies and new oral anticoagulant drugs to prevent future cardiovascular events, it was standard practice to select only those patients with one or more cardiovascular risk factors. Thus although precision medicine will help further refine inclusion/exclusion criteria for clinical trials, such stratification of patients for clinical trials based on individual patient characteristics is not new.

Perhaps one difference now is the further emergence and availability of more "-omic" data, defined by Wikipedia as:

> *"…various disciplines in biology whose names end in the suffix -omics, such as genomics, proteomics, metabolomics, metagenomics, phenomics and transcriptomics. Omics aims at the collective characterization and quantification of pools of biological molecules that translate into the structure, function, and dynamics of an organism or organisms. The related suffix -ome is used to address the objects of study of such fields, such as the genome, proteome or metabolome respectively"*

In addition, we have an ever-increasing capability to measure and use imaging data, biomarkers and PROs to track both disease progression and disease modification from our pharmacological interventions (e.g. PROs could be recorded in a similar way to fitness trackers like Fitbit).

Thus our patient level data is ever increasing both prior to, and during, drug treatment. **Our goal must be to actually utilise this information to deliver better outcomes for patients.**

Recall when we put individual patients outcomes first, the 3 steps we seek to understand in drug development are:

1) Given a patient's individual characteristics, what is the best **initial** drug and dosing regimen?

2) If/when the initial dosing regimen needs to be changed for efficacy and/or safety/tolerability, how best to do this; what is the **best science-based dose titration algorithm?** That is, based on clinical endpoints, biomarkers, imaging and/or patient reported outcomes (PROs), when should the dose be changed, and by how much.

3) Under what circumstances should the dosing regimen be halted?

Thus precision medicine will help at step 1) here. Based on a better understanding of the patient, their disease, and the mechanism of action of the drug, we should be able to better select a drug for a given patient. Indeed, in a paper entitled "Drug Dosing Recommendations for All Patients: A Roadmap for Change", Powell et al. [12] nicely lay out how individual patient characteristics can and should be used to determine an **initial** dose for patients.

However, it is wholly mistaken to equate any **initial** dose recommendation (=guess) as the same as the optimal dose for a patient [13].

Unfortunately we are nowhere near this level of "predictiveness" for PD endpoints. For anyone who thinks that precision medicine will lead to a revolution in how **doses** are determined, I would encourage them to study the history of warfarin, and our best efforts at prospectively determining the right dose of warfarin for each patient [14]. Despite a well-understood PD cascade of coagulation, and the use of individual patient characteristics including genetic polymorphisms, **the determination of the "optimal" dose for a given patient still requires careful dose titrations using on-treatment PD measures;** we simply cannot accurately predict the required dose for a given patient based on their individual patient characteristics alone (as with insulin, anaesthetic agents etc.). Indeed, we have 50+ years of working with much simpler PK data, yet still we have limited capabilities to explain most of the IIV we see across our heterogeneous patients. Rather, we may use these individual patient characteristics to suggest a better **initial** dose, but must then use dose titration to achieve the desired PD response, since our heterogeneous patients are, alas, heterogeneous.

The mantra

"Right drug, at the right dose, to the right patient, at the right time."

has been used as one "vision" of precision medicine. Indeed, we would love to be able to be predictive of, say, a week 16 outcome for a patient based solely on their individual patient characteristics and "-omic" data. However I would provocatively suggest that this information combined would struggle to beat the predictiveness of, for example, their week 4 outcome. I particular like this sentence from Kristensen [15]:

> *"Precision dosing moves beyond the common adjustment of the dose based on body size, demographics factors, renal or hepatic impairment, concomitant medication, etc., and can be guided by observed drug exposure, biomarkers of response or even observed response".*

That is, Personalised Dosing is about using **on-treatment PD data** to intelligently guide dose modifications, and not just using baseline patient characteristics to "magically" determine the right dose for each patient. In addition, even if we could be much better at predicting objective measures of treatment success for any given dose for a particular patient, predicting the subjective assessments of this patient will remain elusive (we cannot tell patients how they should feel!).

Advances across multiple scientific disciplines have always driven drug development, and always will. Further understanding in pathophysiology and pharmacology will continue to drive the design and evaluation of novel drug molecules. Some of these new treatments will be truly fantastic, but I would argue that this has always been the case. **In addition, I am unaware of any precision medicine approach that can tell us the right dose for each patient.** Thus some of the current hype around precision medicine ignores the reality of drug development, the real challenges with poor tolerability/safety, and the central role played by dose. Yes, with a greater understanding of the patient, their disease and the use of "targeted therapies", we will be better placed to select a drug that may have a greater chance of working than other drugs. However most diseases are complex and multifactorial, the patients are complex and heterogeneous, and IIV in PK and PD will remain our enemy.

> **!** Important
>
> Finding the **best** dosing regimen **for each and every patient** will invariably require **Personalised Dosing** using **on-treatment** measures of efficacy and safety/tolerability to guide **appropriate dose titrations**.

Developing an excellent science-based dose titration algorithm may not grab the headlines, but it will matter to patients if it improves their outcomes, so it should matter to us.

Perhaps the next time you read another article eulogising how precision medicine will revolutionise drug development, look to see if they mention how the dose for individual patients will be determined. If not, hopefully this book will still be useful .

11 Dose-Response Modelling; Why We Need Integrated Analyses Across All Doses/Trials

At the end of this chapter, the reader will understand:

- Why we must always "think" and "act" based on integrated analyses across all doses/trials.

- Why we need suitable designs and D-E-R models to fully understand the **shape** of the D-E-R relationships.

- Why interpreting observed outcomes at a specific dose level is naïve, can be misleading, and is never scientifically justified when we have data from multiple dose levels.

- Why investigating narrow dose ranges in few individuals is never an acceptable "dose-ranging" trial.

- That dose-ranging trials need to be prospectively simulated to understand their ability to accurately and precisely quantify the true D-E-R relationships. To simply collect data from a few doses and "hope" for a clear understanding of the D-E-R relationships is unethical.

Fundamentally, if we wish to understand how changes in dose (say from 10 mg to 15 mg) will lead to changes in response, we need to understand the **shape** of the dose-response relationship. I often see D-R data being poorly analysed and interpreted. For example, imagine a "dose-ranging" trial considered three doses of 9 mg, 10 mg and 11 mg. Some hypothetical results are shown in Table 11.1 below (this could be an oncology trial, where Efficacy is Overall Survival and Safety is a Grade 3 toxicity).

Table 11.1: Hypothetical results across three dose levels: 9 mg, 10 mg and 11 mg

Dose	Efficacy	Safety
9 mg	30% (3/10)	10% (1/10)
10 mg	60% (6/10)	10% (1/10)
11 mg	60% (6/10)	40% (4/10)

Given the above data, a naïve person may conclude 10 mg is the "sweet spot", yielding the best efficacy/safety trade-off, since the observed efficacy at 9 mg is inferior to the higher doses, whilst the observed safety at 11 mg is inferior to the lower doses. **It seems that 10 mg has truly the best overall profile, but this is wrong**. We need to think about the cascade from dose to PK to PD; we need to understand the science. This range of doses would yield near identical (and overlapping) exposure ranges across individuals, and hence the true PD effects would be near identical for each of the doses. That is, our best point estimate for Efficacy should be close to 50% (15/30) and close to 20% (6/30) for Safety for **all** doses. Had we actually seen this in a clinical trial, these **observed** results are just one realisation when we "toss a coin" with true response rates for efficacy at some value around 50%, and for safety with a true response rate around 20%; the 10 mg arm just happened to "get lucky", with the 10 individuals at 10 mg yielding, by chance, the most favourable (observed) outcomes.

The above example is extreme in that the dose range chosen (9-11mg) is very narrow, but it clearly highlights one common error seen in drug development, that of naively selecting phase 3 doses based on observed outcomes from small phase 2 trials with very limited dose ranges. In the above example, whilst the pharmaceutical company may expect the 10 mg dose to yield a 60% response for efficacy and 10% response rate for safety in the larger phase 3 trials, in actuality these would be nearer 50% and 20% from the integrated analysis using all dose levels. If we were to imagine that a control arm in phase 3 might yield a 40% response rate, then the pharmaceutical company may design their phase 3 trials assuming a 20% treatment difference (60%-40%) rather than a 10% treatment difference (50%-40%). To achieve 90% power to show a 10% difference required 4 times the sample size compared with a 20% difference. Hence their phase 3 trial would be both underpowered to demonstrate superiority and very unlikely to yield an observed treatment difference anywhere near the 20% that they had expected/hoped for. **The pharmaceutical company has over-estimated the expected response rate in phase 3 by ignoring any form of integrated dose-response modelling and the science that underpins the relationships between dose, exposure and response**. Worse still, had similar results come from a phase 3 trial that investigated closely spaced dose levels, both the pharmaceutical company and regulator may be misled on the actual true response rates across the dose range for both efficacy and safety.

This example shows that it is always dangerous to focus on the observed outcomes for individual dose levels; rather we should always look to combine data across all doses (and trials) to fully understand the D-R relationships. To achieve this, we always need to analysis all data from all doses/trials simultaneously, and utilise a suitable D-R model.

Although experienced drug developers/regulators may think that they would never interpret the above table so naively, consider the following results in Table 11.2

Table 11.2: Hypothetical results across three dose levels: 5 mg, 10 mg and 15 mg

Dose	Efficacy	Safety
5 mg	30% (3/10)	10% (1/10)
10 mg	60% (6/10)	10% (1/10)
15 mg	60% (6/10)	40% (4/10)

Here, I have simply changed the doses from 9-11 mg to 5-15 mg. I make this point based on a recent FDA discussion on Project Optimus, where the dosing regimen for idelalisib (brand name Zydelig) was discussed. The discussion centered on how the approved 150 mg bid dose was now considered too high. The discussant then showed a table with efficacy and safety results based on doses from 50 mg qd to 350 mg bid. The results were based on a small N for each dose regimen, and hence were not very dissimilar to the above. Based on the observed safety data at 100 mg bid looking **numerically much better** than that at 150 mg bid, one participant suggested that 100 mg bid could be a **much** better/safer (optimal?) dose than 150 mg bid. Is this suggestion sound?

From the FDA label, we can better understand some key information about the relationship between dose and exposure for idelalisib (my **emphasis**):

> *"Idelalisib exposure increased in a less than **dose-proportional** manner over a dose range of 50 mg to 350 mg twice daily in the fasted state."*

When we have dose proportionality, we expect the exposures at 100 mg bid to be, on average, 67% of those at 150 mg bid (67% = 100/150). However the above suggests that the exposures at the 100 mg bid are likely to be **closer** to the exposures at 150 mg bid. The IIV in clearance for idelalisib (directly related to IIV in exposure) is quoted as 38.6% [16], the typical exposure (measured by AUC) is quoted as 10010 ng.h/mL, and we can (conservatively) approximate that the exposures at 100 mg bid are likely to be nearer 75% of the exposures at 150 mg bid (they may be even closer, and the IIV even higher, but these ballpark figures will suffice to make the point here). Figure 11.1 shows the distributions of exposures (as measured by AUC) for 150 mg bid in the two top panels (these are identical except for random sampling differences across 10000 simulated patients). The bottom panels show the distribution of exposures when we use the Not Proportional estimate (75%) and the Dose Proportional (67%) value for 100 mg bid.

Under both the Not Proportional and Proportional scenarios, the distributions of exposures significantly overlap across the two dose levels. Recall that when the distributions of PK exposures overlap, the PD effects **will** be similar (even if we tried here to foolishly use a step function for the exposure response relationship). Thus purely on understanding the IIV
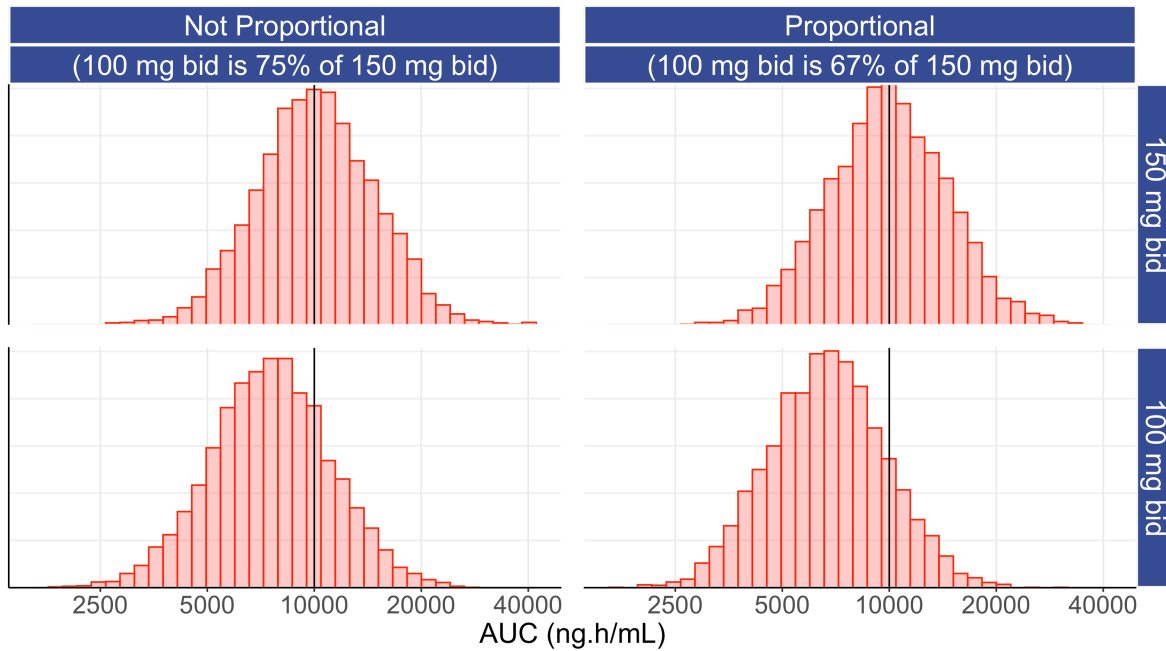
Figure 11.1: Distribution of exposures for idelalisib for 150 mg bid (top) and 100 mg bid (bottom) assuming a Not Proportional (left) and Dose Proportional (right) relationship between the two doses. The median AUC at 150 mg bid is shown as a vertical reference line (10100 ng.h/mL)

for a drug, how close the two dose levels are, and by understanding that exposure response relationships for PD do not follow step functions, we can make informed observations about the relative effects of these two doses.

Returning to the above table for our fictitious drug/trial, our 15 mg dose is 50% higher than 10 mg (i.e. similar to the ratio of 150 mg bid to 100 mg bid with idelalisib). Thus for our fictitious drug (or idelalisib), is it possible that the D-E-R is very steep, and 10 mg (100 mg bid) is quite safe, when 15 mg (150 mg bid) is not?

In short, no, science tells us that this is not possible.

If the 150 mg bid dose of idelalisib was found in later phase 3 trials to have a poor safety profile, the safety profile at 100 mg bid would not be "clean".

D-E-R relationships do not follow "step functions" where we suddenly go from no effect to a much larger effect.

In addition, when doses are close with overlapping exposure ranges, we can be confident that the safety profiles of the two doses will not be very different. Thus although idelalisib 100 mg bid may well have a **better** safety profile compared with 150 mg bid, we can only accurately quantify these differences from well conducted trials using very wide dose ranges, sufficiently large sample sizes and appropriate D-E-R modelling incorporating all doses/trials. Simply"eye-balling" the observed data for each dose is neither a scientific nor reliable method for any form of (accurate) dose selection.

To conclude the discussion around the results for our fictitious drug trial, we cannot accurately or precisely determine the D-R for either efficacy or safety from such a weak design; the dose range is far too narrow and the sample size is far too small. When faced with such limited data from such an awful trial design, companies /teams /regulators /individuals may be compelled to simply select the dose regimen that looks numerically "best", but I see this as analogous to choosing black or red on the roulette wheel based on the result of the last spin; it is painfully unscientific and unsound.

11.1 Example of a Weak Dose-Response Design and Analysis

*A real example of the challenges with interpreting data from small trials and narrow dose ranges in oncology is discussed in this section. It includes some observations on **Project Optimus**, the FDA initiative to reform the dose optimization and dose selection paradigm in oncology. To be useful, the material is necessarily detailed in places, and hence the reader may wish to skip this section.*

The following example shows the willingness to interpret results for individual doses levels can be misleading both to the drug company and the regulator. As stated above, we should both "think" and analysis all data from all doses/trials simultaneously via suitable D-E-R

models. This provides us with the most accurate and honest understanding of the true D-E-R relationships, and hence the best evidence base to make informed decisions. However consider the following trial called DESTINY-LUNG02 (NCT04644237), where two dose levels of trastuzumab deruxtecan (T-DXd) were investigated for the treatment of HER2-mutated Metastatic Non-Small Cell Lung Cancers (NSCLC). The two dose levels were 5.4 mg/kg and 6.4 mg/kg, and the FDA has granted accelerated approval. Selected text is shown below from the following source:

https://ascopost.com/issues/september-10-2022/fda-grants-accelerated-approval-to-trastuzumab-deruxtecan-for-her2-mutant-nsclc/

> *"T-DXd was evaluated at a 6.4 mg/kg dose across multiple trials and at a 5.4 mg/kg dose in a randomized dose-finding trial. Response rates were consistent across dose levels. Increased rates of interstitial lung disease/pneumonitis were observed at the higher dose. The efficacy results of the approved recommended dose of 5.4 mg/kg given intravenously every 3 weeks are described below."*

> *"Of the 52 patients in the primary efficacy population…"*

> *"The confirmed objective response rate was 58% (95% confidence interval [CI] = 43%–71%) "*

> *"This application used advice from the FDA Oncology Center of Excellence (OCE) Project Optimus to conduct a dose-randomization study, which led to a lower dose being approved. For more information regarding the OCE's efforts to modernize dose selection for oncology products, refer to Project Optimus."*

A parallel trial called DESTINY-LUNG01 [NCT03505710] only considered the 6.4 mg/kg dose in 91 patients, it reported a high rate (46%) of grade 3 or higher drug related adverse events [17]. It stated:

> *"The safety profile was generally consistent with those from previous studies; grade 3 or higher drug-related adverse events occurred in 46% of patients, the most common event being neutropenia (in 19%)"*

To be clear, I am neither an expert in NSCLC nor aware of the unmet need in this patient population. However as a drug developer, there is so much here that looks highly questionable from a dosing justification perspective, and how the limited data collected is being interpreted. The response rate in DESTINY-LUNG02 of 58% would seem to equate with 30/52 patients achieving the objective response, thus we have **very few patients**. The text proposes that the "response rates were consistent across dose levels", but the sample sizes are far too small to draw any meaningful conclusions (we should not be guessing!). It also states that the lower dose (5.4 mg/kg) had a (numerically) better safety profile and, based on this, the lower dose was approved by the FDA. **The label will now report the results for the 5.4 mg/kg dose**. With such small N, and two doses levels that are so close to each other, such interpretations of the relative efficacy and safety of these two doses is highly speculative at best, and simply

pure guesswork at worst. Whilst it is perfectly reasonable that the efficacy profile at the two doses were **numerically** in the same ballpark, and that the safety for 5.4 mg/kg did look **numerically** better than the 6.4 mg/kg dose, there is a substantial risk we are trying to make informed decisions around the dose regimen based on weak data that is incapable of being used for such decisions. **Why is there a risk to the regulator (and patients)?** Imagine I am an unscrupulous drug developer, and I want my label to look as good as possible. Under the cloak of supporting Project Optimus, I use 4 doses that are all very close (e.g. 5.4 mg/kg, 5.7 mg/kg, 6.0 mg/kg and 6.4 mg/kg). By having such similar doses and small N, I am essentially maximising my chances that one looks "optimal" by chance, and better than it would truly be. I can get a label based on my "lucky" regimen, although a truly well designed trial (with a very wide dose range) and integrated D-E-R analysis would put these dose levels at broadly similar levels of efficacy/safety. It is analogous to using 5 different shades of blue pills into a trial, and then subsequently seeking approval for the darkest blue pill, since the **observed** data for that shade of blue happened to look best when compared to the other shade of blue pills.

As an aside, everything above has focussed on a fixed-dose regimen for trastuzumab deruxtecan that is "personalised" using only body weight. In the DESTINY-LUNG01 trial, a total of 88 patients (97%) had drug related adverse events, with 31 patients (34%) having drug-related adverse events that led to dose reduction; 23 patients (25%) discontinued treatment. Drug-related interstitial lung disease occurred in 24 patients (26%), with Grade 1-5 neutropenia (low neutrophil counts (a type of white blood)) being experienced by 32 patients (35%) and Grade 3-5 in 17 patients (19%). A very long list of additional adverse events was reported; **this dosing regimen is extremely tough on the patients**. The percentage change in tumour dynamics was also reported for each patient, and is reproduced from Li [17] in the Figure 11.2 below.
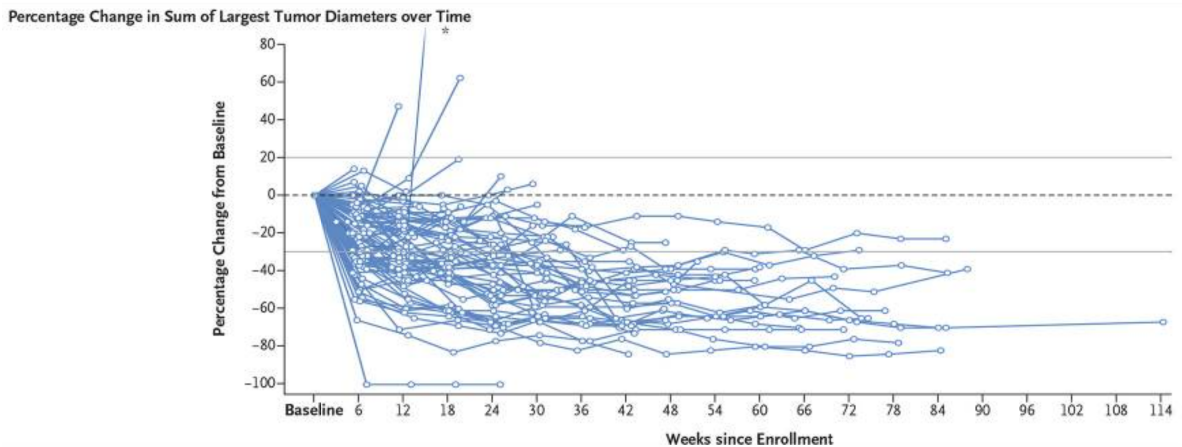

Figure 11.2: The percentage change from baseline in the sum of the largest diameters of measurable tumors from baseline over time. The asterisk indicates a patient outlier with an increase of 236% in tumor diameter from baseline at week 18.

Although these patients all received the same 6.4 mg/kg dose, this figure clearly demonstrates the IIV between patients in a relevant PD endpoint. Following a more comprehensive analysis that looked closely at, for example, the interrelationship between overall survival, tumour dynamics, changes in neutrophil counts and pharmacokinetics, would it not be possible to construct a **science-based dose titration algorithm** based on, for example, accruing on-treatment data for changes in neutrophil counts, PROs and tumour dynamics? For example, the 17 patients who experienced grade 3-5 neutropenia would not have experienced such severe neutropenia had they started on a lower dose; in our urgency to deliver a very high dose, we are exposing patients to significant toxicities. **Is it not important to investigate whether a significant reduction in the tumour diameters can be achieved with lower doses for some patients?** Many PKPD models to describe neutrophil counts and predict neutropenia have been developed, so can we not initially titrate the dose to ensure that, at worst, only grade 1 (or 2) neutropenia are observed? Given 25% of patients discontinued treatment in this trial, and the very high number of severe toxicities, are we not ethically compelled to consider such trials? I absolutely think we should be doing better here; intelligently titrating the dose (say over the first 3-18 months) is a small logistical price to pay if we truly care about patients and their outcomes.

In summary, the development of trastuzumab deruxtecan has involved multiple indications beyond NSCLC, so the appropriateness of the dosing regimen proposed in this indication will naturally be augmented by other data. However there is no expert in D-E-R modelling who would consider a dose "range" from 5.4 mg/kg to 6.4 mg/kg as being capable of generating meaningful data for informed D-E-R analysis. The adage "**Garbage in, Garbage out**" is unfortunately how we need to view such "dose ranging" trials. Before running any such trial, it is essential that clinical trial simulation be used to assess whether the data that will be generated will provide any meaningful quantification of the D-E-R relationships. If such an exercise was performed using doses like 5.4 mg/kg and 6.4 mg/kg in 25-50 patients per arm, **I know** the virtual trials would show a random collection of outcomes, with 5.4 mg/kg sometimes looking superior to the 6.4 mg/kg arm, with other virtual trials showing the exact opposite. **I would argue it is unethical to utilise such a poor design; more informative designs should have been used. Would you agree?**

If we truly wish to understand D-E-R relationships, we need appropriate designs and appropriate analysis. In areas outside of oncology, well-designed trials that investigate very wide dose ranges have been successfully used to characterise D-E-R relationships across multiple efficacy and safety endpoints [6] . If Project Optimus is to be truly successful, it must ensure very wide dose range are investigated, and ensure the data from all doses be combined with suitable D-E-R models. Simply advocating a few, closely spaced, dose levels is neither sufficiently informative nor acceptable from either a scientific or patient perspective; we must do much better.

12 Introducing The Most Important Dose-Response Model

At the end of this chapter, the reader will understand:

- Based on 70+ years of accumulated knowledge in pharmacology, that it is appropriate to assume that the effects at different doses are related of each other.

- That to quantify dose-response relationships, we should (initially) use the sigmoidal Emax model.

- The definitions and meanings of each of the 4 parameters of the sigmoidal Emax model (E_0, E_{max}, ED50 and Hill coefficient).

- The parameters values for the Hill coefficient (=steepness) of D-R relationships that are typically seen in drug development.

- Why doses that are closely spaced (e.g. 1 mg and 2 mg) can yield similar results.

- Why D-R trial designs need very wide (think 10-100 fold) dose ranges.

- Why "simpler" dose-response models (linear, log-linear, quadratic etc.) are inappropriate, and should never be considered.

This chapter is essential reading for anyone wishing to understand and interpret dose-response data in drug development.

At the core of understanding dose-responses in drug development is the need to understand how the effects at different doses relate to each other. We need to understand what is credible, and what is not credible, based on the 70+ years of accumulated knowledge in pharmacology. We have two choices:

1. Assume the effects at different doses are **unrelated** of each other, and hence analyse the observed effects at each dose without consideration of the effects at other doses.

2. Assume the effects at different doses are **related** of each other, and hence analyse the observed effects across all dose simultaneously.

1) is wrong; it is unscientific. 2) is correct, but implicitly requires the use of a suitable D-R model to "link" the doses together. Consider the following simple results shown in Table 12.1

Table 12.1: Hypothetical results for placebo and two dose levels for a safety endpoint

Dose	Safety
Placebo	10% (1/10)
10 mg	10% (1/10)
20 mg	40% (4/10)

It is always tempting to interpret the observed results at each dose as exact (i.e. without error) values, and indeed often D-R graphs simply "join the dots" across the average responses at each dose. In the above example, only 2 of the following 3 statements could be **true**:

1) Neither dose has an increased safety risk relative to placebo

2) Both doses have an increased safety risk relative to placebo

3) Only the 2 mg dose has an increase safety risk relative to placebo

Statement 1) is potentially true; neither of the two doses truly affects this safety endpoint, and hence the observed data are random samples from some common effect (e.g. the true rate is 20% for all arms, and the 20 mg arm looked, by chance, worst). **Statement 2) is potentially true**; both of the two doses affect this safety endpoint, and our best estimate of the true effect would be based on an integrated analysis across all doses (although using just two doses, like in the above, would be painfully limited (=> need a better design!)). **Statement 3) is false.** Here we are talking about the truth, not the observed outcomes. Based on our earlier exposition linking dose to PK to PD, and understanding IIV, we never see "step function" dose-responses (recall a "step function" is where the effect increases (like a step) from no effect to some new effect at some (magical!) place on the dose or exposure scale). Pharmacology tells us this does not happen as we move from 10 mg to 20 mg.

I hope the above has sufficiently explained that naively "joining the dots" for the average responses for each dose level is unscientific. **To assume the effects at different doses are unrelated to each other is inconsistent with our understanding of pharmacology over the last 70+ years.** In contrast, to recognise that doses are related to each other makes perfect sense, but we will require a mechanism (model) to allow us to analyses all doses simultaneously to make informed decisions about the true, underlying D-R relationships. Fortunately, we know how to do this.

There are many ways of linking responses to doses, but there is one D-R model that stands supreme above all others. This model is called the **4 parameter sigmoidal Emax model**, and is defined as:

$$E = \ E_0 + \frac{E_{\max} \bullet Dose^\gamma}{Dose^\gamma + ED50^\gamma}$$

Here E is the response, E_0 is the response associated with Dose = 0 (e.g. a placebo response rate), $E_{\max}$ is the maximum drug effect, ED50 is the dose required to yield 50% of the $E_{\max}$, and γ is the Hill coefficient, which defines the steepness of the dose-response (under an alternative parameterisation, this model is also known as the 4 parameter logistic model).

The influence of each parameter on the shape of the dose-response is shown in Figure 12.1.

These figures show how the E_0 parameter shifts the whole D-R up or down, how the $E_{\max}$ parameter changes the maximum drug effect, how the ED50 parameter shifts the D-R to the left or right, and how the Hill parameter captures the steepness of the D-R. In addition to these parameters having meaningful interpretations, the combination of these 4 parameters provides sufficient flexibility to describe a very wide range of D-R relationships, and this model has consistently been an excellent foundation for my D-E-R modelling over the last 28 years.

This model is further illustrated in Figure 12.3 for an E_0 of 0, an $E_{\max}$ of 100, an ED50 of 10 mg, and Hill coefficients of 0.5, 1, and 2 using a logarithmic scale (top) and an untransformed scale (bottom) for dose. Responses for dose levels that are 2, 4 and 10 times lower/higher that the ED50 are shown as text. For simplicity, the parameters chosen here yield a 0-100 range for our response, but ordinarily they will reflect the placebo response (E_0) and drug specific parameters ($E_{\max}$, Hill, ED50) for the particular endpoint/drug. Because different E_0 and $E_{\max}$ parameter values simply rescale the response axis and a different ED50 value would simply shift the dose-response left or right, we can freely generalise this illustrative case to any D-R.

Of note, although the logarithmic scale (top) is best for understanding the "shape" of the D-R, it requires a "home" for a dose of 0 (i.e. placebo). In the top panel, it has been placed substantially to the left of the lowest dose. Note the symmetry in the dose-response in the logarithm plots either side of the ED50. In contrast, the untransformed scale (bottom) accommodates a dose of 0, but it is overly dominated by the highest dose, with the lower dose levels heavily compressed to the left of the dose scale (this problem is even worse when using exposure, since the highest exposure at the highest dose will dominate the drug exposure scale). As such, the 3 D-R models on the untransformed scale do not show the clear difference in "shape" as compared with the logarithmic scale. I would always recommend that both figures be shown.

My experiences of D-E-R modelling guides me to view a Hill coefficient of 1 as being a good "ballpark" figure for a "typical" D-R, although we should never fix this parameter at the estimation stage (since, like the ED50, it is a feature of the drug/endpoint that we wish to quantify precisely). It can be seen that the 10 mg dose yields an effect of 50 that rises to 67

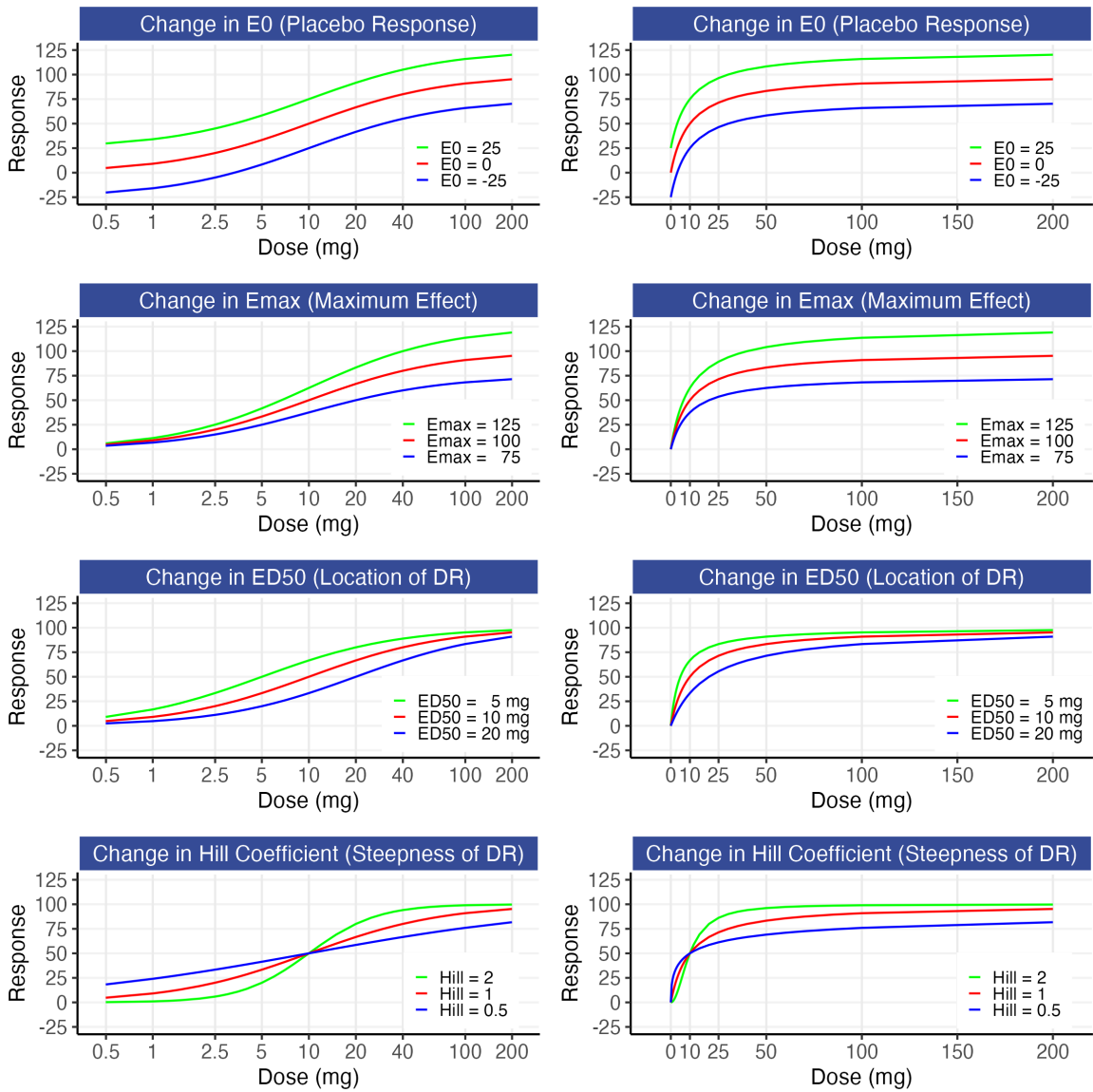Figure 12.1: Illustration of the influence of each parameter on the shape of the dose-response for dose on the logarithmic scale (left) and untransformed scale (right). Unless stated otherwise, E0 = 0, Emax = 100, ED50 = 10 mg, Hill = 1.
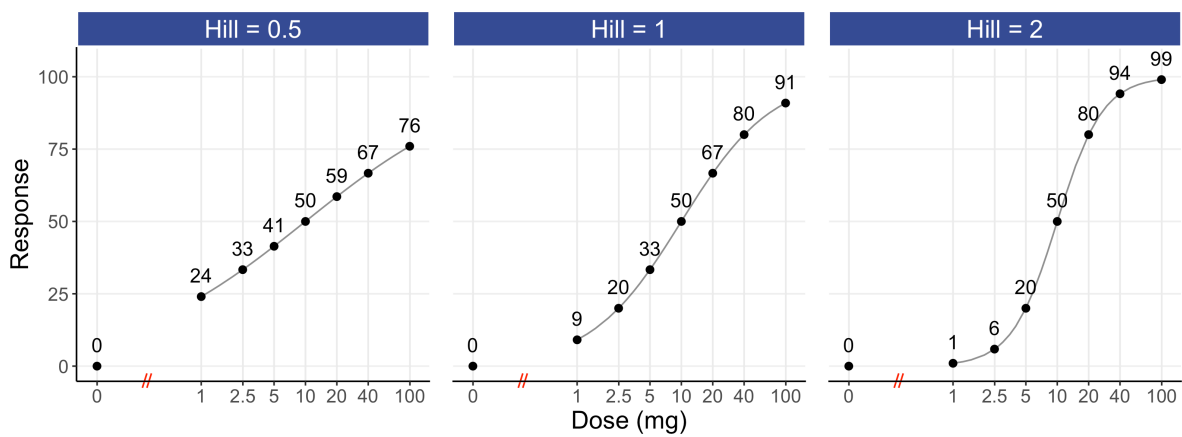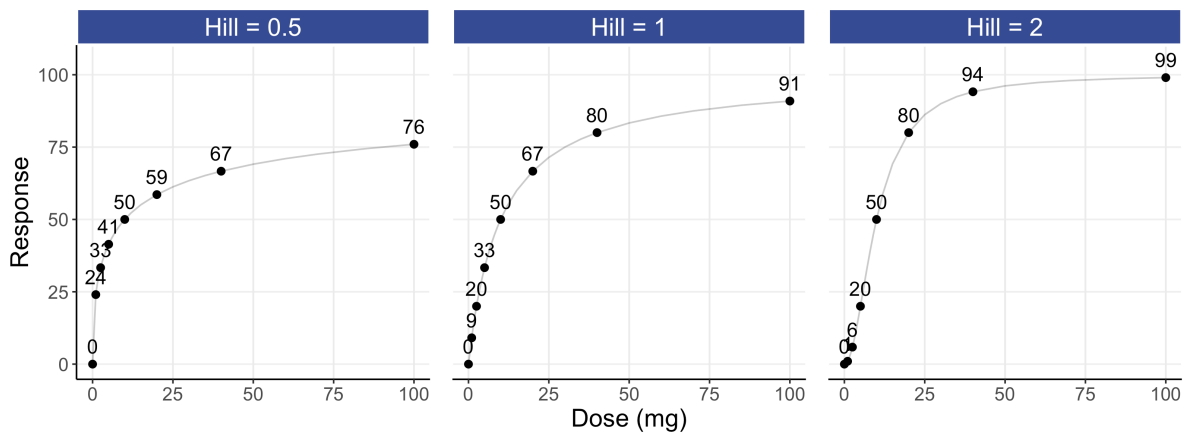
Figure 12.2: .

Figure 12.3: Illustration of the steepness of plausible D-R relationships on the logarithmic scale (top) and untransformed scale (bottom) for dose

at 20 mg. **Thus around the middle of a "typical" D-R, a doubling of dose yields a relatively modest increase in effect**. Please take a moment to reread this last sentence; a two-fold increase in dose will **not** generally lead to a large increase in effect. For doses that are higher than the ED50, a doubling of dose (e.g. 20 mg to 40 mg) will yield a smaller relative increase (i.e. as we near the "plateau" of the D-R curve, a doubling of dose has less marginal benefit). For most drugs, therapeutic doses are more likely to be **above** the ED50 for efficacy endpoints, so although a doubling of the dose will lead to a higher response rate, the incremental benefit may not be as much as we might hope. In contrast, therapeutic doses are more likely to be **below** the ED50 for most tolerability/safety endpoints, as we are generally in the bottom part of the D-R curve for these endpoints (e.g. we might expect 10-40% of patients to be experience neutropenia or nausea at the therapeutic doses, and not 60-90% etc.). At the lower part of the dose-response, a doubling of dose will generally lead to a greater "jump" in response (e.g. from 20 at 2.5 mg to 33 at 5 mg (a 65% relative increase). Thus when we titrate drugs it is generally the safety considerations, and not efficacy, that will guide the magnitude and timing of dose titrations. In particular, to ensure we can most efficiently find the right dose for each individual, we wish to understand IIV in the location (ED50i) and steepness of the safety curves within individuals.

If we review the shape of the D-R for a Hill coefficient of 1, we can see how "dose-ranging" trials that have narrow dose ranges (e.g. 4 fold) are going to very poorly estimate the true shape of the D-R. For example, even if the doses chosen perfectly spanned the ED50 (i.e. giving the largest difference between the lowest and highest doses), then the 5 mg response (33), the 10 mg response (50) and the 20 mg response (67) only cover a 2 fold range in the response domain. With noisy/variable endpoints and small N (as often used in phase 2), the observed (estimated) D-R relationship from such a poor design will be painfully inaccurate and can easily be misleading. I would encourage anyone considering such a trial design to perform a simulation-estimation across 1000 virtual trials for a range of sample sizes. You will discover enormous variation in the "observed" D-R across the 1000 virtual trials, with only enormous sample sizes being sufficient to recover the "true" D-R used in the simulation.

To fully characterise the D-R, we can think about the ED10 dose (the dose required to yield 10% of Emax) and the ED90 dose (the dose required to yield 90% of the Emax) as ballpark reference points that span a wide range of the D-R. For a Hill coefficient of 1, the ED10 is 1.11 mg, and the ED90 is 90 mg. This equates to an **81-fold** range (90 mg / 1.11 mg). If we wished to only focus on ED50 and above, this equates to a **9-fold** range in doses (90 mg / 10 mg). **I hope this brings into sharp focus just how badly most "dose ranging" trials are designed; there are invariably far too few dose levels spanning a far too narrow dose range.**

> ❗ Important
>
> As a broad guiding principle, the above highlights than we should aim to have 10-100 fold dose ranges in the "interesting" part of the D-E-R relationships throughout drug

development. With sufficient N, this should enable us to precisely quantify how both efficacy and safety endpoints change across the whole dose range investigated.

As an experienced/old analyst, I have often been asked to look at poorly designed "dose-ranging" trials, and I am always reminded of this wonderfully acerbic quote by RA Fisher:

> *"To call in the statistician after the experiment is done may be no more than asking him to perform a post-mortem examination: he may be able to say what the experiment died of."*
>
> *Presidential Address to the First Indian Statistical Congress, Sankhya 4, 14-17, 1938.*

A skilled analyst can fix a poor analysis, but they cannot fix a poor design, and for D-R trials, this is when the sponsor only investigates a very narrow dose range. Indeed, I would argue it is unethical to run "dose-ranging" trials when they are incapable of meaningfully capturing the true D-R; we are wasting the time, effort and altruism of the patients in these trials. Such poorly run trials invariably leads to the sponsor simply picking the "best looking" dose based on the observed outcomes at each dose, without appreciating that these observed outcomes are always "noisy". In these cases, both the sponsor and regulator may easily be misled (as discussed in the previous chapter).

The choice of Hill coefficients shown in Figure 12.3 covers the range seen in drug development, with a Hill coefficient of 0.5 representing a very shallow D-R, and 2 representing a very steep D-R [18]. In my experience across multiple endpoints and therapeutic areas, I have never seen (from well designed trials) Hill coefficients outside of the 0.6-1.5 range, so the values of 0.5 and 2 can be viewed as extreme cases that provide a guide for us to better understand the "range" we might expect to see for any endpoint/drug combination. When I once asked Lewis Sheiner a question, he replied "Well, we always know something". His point was that we have copious amounts of prior experience/knowledge that we can draw upon to augment our understanding/analyses. Recall our very weak trial results, now shown in Table 12.2 with a "model predicted" column.

Table 12.2: Hypothetical results for placebo and two dose levels for a safety endpoint along with the model predicted estimates

Dose	Safety	Model Predicted
Placebo	10% (1/10)	10.00%
10 mg	10% (1/10)	10.00%
20 mg	40% (4/10)	39.99%

As previously discussed, this is a weak design. If an analyst tried to fit the sigmoidal Emax model here, the estimation algorithm would stop at one possible solution (there is no unique

solution here). For example, the "final" parameter estimates might have $E_0 = 10\%$, $E_{max} = 30\%$, ED50 = 15 mg, and a Hill-coefficient = 20. This combination of parameters yields the "Model Predicted" values shown in the table. These predicted responses at each dose are very close to the observed responses at each dose level, since this combination of parameters yields a "step function" type dose-response that seemingly "fits" the observed data very well. Although is may be appealing to conduct each analysis in a vacuum (under the guise of "objectivity"), the above example shows how this can lead to nonsense. We know that Hill coefficients of 20 are impossible for clinical endpoints used in drug development; hence the estimated model parameters and estimated D-R here is, for sure, not credible.

A more refined analysis could restrict the Hill coefficient to the 0.5 to 2 range and incorporate additional information on the potential magnitude of the Emax parameter. Although this would lead to a more credible set of plausible D-R relationships that excludes nonsense (such as a Hill coefficient of 20), fundamentally the weakness of the design is crushing (an E-R analysis would likely be better than D-R here, but we are still "clutching at straws"). As always, the right answers can only be determined from sufficient data using the right design.

The estimation of the Hill coefficient is an integral component of D-R modelling, since it quantifies how the response changes across key dose levels. For example, consider the ratio of the response at 20 mg (2*ED50) to 5 mg (0.5*ED50). For Hill coefficients of 0.5, 1 and 2, this ratio is 1.41, 2 and 4 respectively. Thus a 4-fold change in dose either side of the ED50 leads to a very modest increase in response (41% higher for Hill = 0.5) to a substantial increase in response (300% higher for Hill=2). Since most drugs are dosed well above the ED50 for efficacy endpoints, we can also compare the increase in response between two higher doses, for example 40 mg (4*ED50) compared to 20 mg (2*ED50). In this case the ratios of responses are 1.14, 1.20 and 1.18 for Hill coefficients of 0.5, 1 and 2 respectively. For higher doses still, for example 100 mg (10*ED50) compared to 50 mg (5*ED50), the ratios are 1.10, 1.09 and 1.03 for Hill coefficients of 0.5, 1 and 2 respectively. **Thus the higher the dose above the ED50, the less additional benefit is observed with "pushing the dose"**. In contrast, the dose levels of interest for safety/tolerability endpoints will generally be below the ED50, and hence a doubling of dose will lead to proportionally greater increases in the (unfavourable) responses. For example, going from 1 mg (ED50/10) to 2 mg (ED50/5), the ratio of the responses increase by 1.29, 1.83 and 3.88 for Hill coefficients of 0.5, 1 and 2 respectively. **Thus when we consider the role of optimal dose ranges or an optimal titration strategy, the precise and accurate understanding of these D-R parameters will be critical**. In the case when we have a shallow (Hill=0.5) D-R relationship, a simply two fold increase in dose may have only a very modest change in both efficacy and safety/tolerability, whereas when we have steep (Hill=2) D-R relationship, a simply two fold increase in dose can have a significant change when the dose is towards the lower or middle part of the D-R range, but minimal effect at the higher dose levels.

Thus we can **only** discuss the role of dose when we fully understand the **shape of the dose-response** for both efficacy and safety/tolerability in terms of the location (ED50), maximum effects (Emax) AND the steepness of the dose-response (Hill coefficient).

I would anticipate that at least 95% of dose-responses seen in drug development could be adequately described using the sigmoidal Emax model as the foundation for the development of an acceptable D-R model. To achieve this though, we need appropriate trial designs (very wide dose ranges and large sample sizes). Occasionally very high doses can lead to attenuation in the response (i.e. non-monoticity in the dose-response), and in rare circumstances asymmetry in dose-response (on the logarithmic scale) may be evident (suggesting a more complex model, like the Richard's model, may be needed). However in general this model is an excellent starting point for D-R modelling that can be augmented when necessary. Note when very few, closely spaced, doses are investigated using a noisy endpoint with few patients, it is not at all unusual to see the observed dose response appear "odd" (e.g. linear, quadratic, "umbrella", "n", "j" shaped, non-monotonic etc.), but this is a reflection of the very poor trial design, with any post-hoc exercise in D-R modelling unlikely be very helpful (the "Garbage-In, Garbage Out" scenario).

The above 4 parameter sigmoidal Emax model can also be used for exposure-response modelling, by simply substituting either the observed (or predicted) concentration (C) or an exposure measure (such as C_{ave}, C_{trough}, C_{max}) for Dose, and EC50 for ED50, such as:

$$E = E_0 + \frac{E_{\max} \bullet C_{ave}{}^{\gamma}}{C_{ave}{}^{\gamma} + EC50^{\gamma}}$$

Here EC50 is the average concentration that leads to 50% of the maximum effect. Although the model is very similar to that using Dose, moving between Dose, C_{ave} and C will subtly change the model parameters, as will be discussed later.

Importantly, when responses are observed only for a single dose level, we clearly have no idea of the shape of the dose-response relationship, and a very poorly informed understanding of the exposure-response relationship.

Although the exposure range across patients may be very wide, it can be misleading to simply assume the observed exposure-response in wholly accurate since it is dose, not exposure, which is randomised. For example, the higher exposures may be observed in patients with renal impairment who might be, on average, typically older and sicker that other patients. Thus simply ignoring this can lead to an erroneous understanding of the true exposure-response relationship; it may be flatter or steeper than what is observed.

Put simply then, if only data from a single dose level is available, our ability to characterise the steepness/shallowness of the D-E-R for efficacy and safety/tolerability is highly compromised.

What happens if an elderly patient accidently takes two pills instead of one? From a safety perspective, absence of such knowledge should be a major concern for all, and should be seen as unacceptable in modern drug development; we always need multiple doses across wide dose ranges, even when an E-R model will be used.

Finally, it is worth briefly mentioning some of the weak alternatives to the sigmoidal Emax model I see used as "dose-response" models. These include the linear models, log-linear models, quadratic/umbrella models, and simple Emax models (where the Hill coefficient is fixed to 1). **These are, universally, awful**. Being unkind, there use indicates insufficient training and experience in D-E-R modelling to recognise the limitations with these models from both a pharmacological and statistical perspective. In addition, methodologies (such as MCP-MOD) which attempt to encompass such models within the framework of "selecting" or "model averaging" across such models are equally misguided. I generally refrain from spending time critiquing weak statistical methods in drug development; the list is rather long . However I do hope to add a chapter/appendix to this book to specifically cover why these alternatives are so inappropriate as D-E-R models. This is because, as an inexperienced analyst, I made some of these same mistakes myself, so am acutely aware of the lack of foundational (any?) training material on this important topic. This book should help others avoid some on these errors.

---

13 Population and Individual Dose-Exposure-Responses, Therapeutic Windows and Maximum Tolerated Doses

At the end of this chapter, the reader will understand:

- That we need precise language when we discuss D-E-R, therapeutic windows and maximum tolerated doses; **Population** and **Individual** levels are very different.

- That knowing the shape of the **Population** D-E-R does not tell us about the shape of **Individual** D-E-R relationships.

- Why **Individual** optimal doses will be markedly different to any **Population** optimal dose; no one dose is optimal.

- Extensive experiences with anaesthetic agents, warfarin and insulin show us that to achieve the same response, the optimal doses for patients span very wide dose ranges (think 10-50 fold).

It is common to hear discussions on the "dose response" of a drug, but to truly understand and communicate what we mean by this term, we need to be much more exact in our language. This is because **Personalised Dosing** requires us to think and understand "dose response" at the **Individual** level, not the **Population** level.

Fundamentally, when we discuss the D-E-R relationship, a key distinction needs to be made between the **Population** D-E-R relationship and **Individual** D-E-R relationships. Figure 13.1 shows the relationship between the **Population** D-E-R and **Individual** D-E-R relationships for a hypothetical drug (for simplicity, just the D-R is shown). The efficacy, tolerability and safety responses are plotted versus dose for 100 individuals (the gray lines); **importantly, Individuals do not all have the same D-R, but have their own D-R** (this is the IIV in D-R discussed previously). The final panel of the figure shows a simple clinical utility index
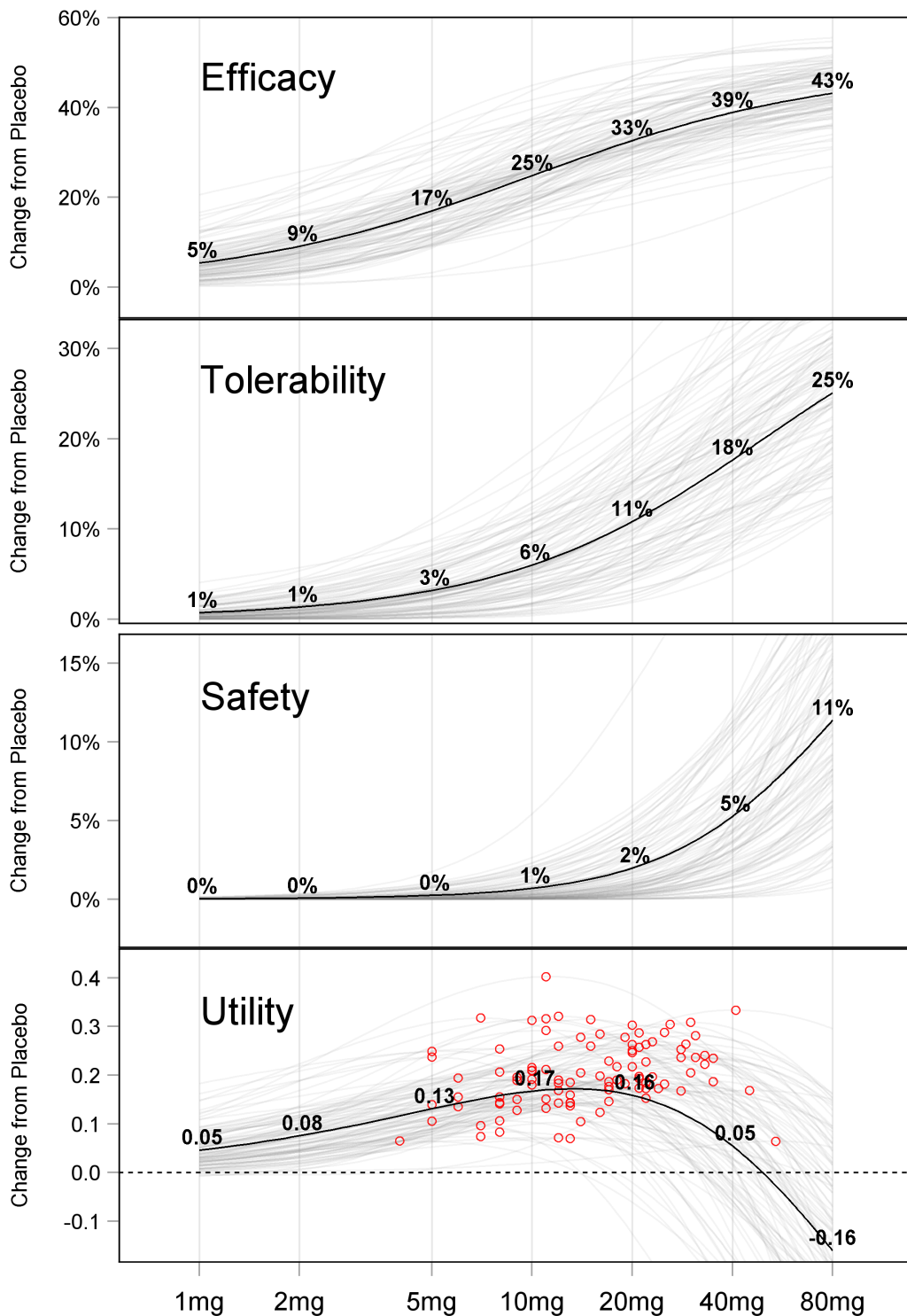
Figure 13.1: The Population D-R relationship (bold line) and 100 individual D-R relationships (gray lines) for an Efficacy, Tolerability and Safety endpoint for a hypothetical drug. The Population- D-R is simply the mean of the Individual D-R relationships. The final panel (Utility) shows a very simple clinical utility index, where 'Utility score = Efficacy – Tolerability - 3*Safety' (e.g. the relative weighting is 1:1:3 for the three endpoints). The last panel highlights the maxima of the individual utility curves (red circles), showing that they are widely distributed over the dose range. No one dose is optimal for all patients

(CUI) that balances the "benefits" in efficacy with increasing dose with the "harms" in terms of tolerability/safety with increasing dose.

The Population D-R (the bold line in the figure above) simply joins together the average of the individual responses for each dose. Since the individual D-R relationships with dose are non-linear, the Population D-R is not even the dose response for a "typical" individual; individuals can have very different individual D-R relationships compared to the Population D-R. Thus an increase in dose at the individual level will not "move patients along" the Population D-R curve, rather it will move them along **their** curve.

> **!** Important
>
> **The Population D-R relationship does not tell us anything about the underlying D-R relationships for individual patients.**

In contrast, the individual D-R relationships (shown in gray) are exactly what we seek to understand, and hence determine the optimal dose for each individual (and indeed the optimal dose titration algorithm). The final panel of the figure, labelled Utility, highlights with red circles the dose for each individual that maximises their individual utility (i.e. the best dose for each individual based on a simple CUI). Note how the **Population** utility curve has a maximum at 10 mg, but that the individual "optimal" doses in this simple example range from below 5 mg to greater than 40 mg, spanning more than a 10-fold range across the 100 individuals.

The fact that the **Population** D-R does not tell us about **Individual** D-R relationships is demonstrated in Figure 13.2.

In the top 3 panels, the Population D-R under 3 different scenarios are shown alongside 100 random individual D-R curves. The three population D-R curves are shown superimposed in the bottom panel. Thus when we only observe the **Population** D-R, we cannot infer the shape of the **Individual** D-R relationships. This highlights the primary deficiency with **Population** D-R relationships; if we wish to understand **Individual** D-R relationships, we need different trial designs.

Thus in drug trials where each individual only contributes data from a **single** fixed dose regimen (or exposure measure), the data generated across individuals/doses can **only** determine the **Population** D-R. Most phase 2 "dose-ranging" trials use this crude parallel group design, where different cohorts of individuals receive different doses levels (e.g. placebo and 5 mg, 10 mg, 20 mg, 40 mg and 80 mg); there is no within-individual dose titration. The resulting D-E-R relationships are therefore just **Population** D-R relationships, and hence any decision on the (optimal) "one-size-fits-all" dose is wholly ignorant to the underlying **Individual** D-R relationships. Stated alternatively, different **Individual** D-R relationships can yield the same **Population** D-R relationships, and hence knowledge of the **Population** D-R does not provide the required information to enable us to find the right dose for each individual.
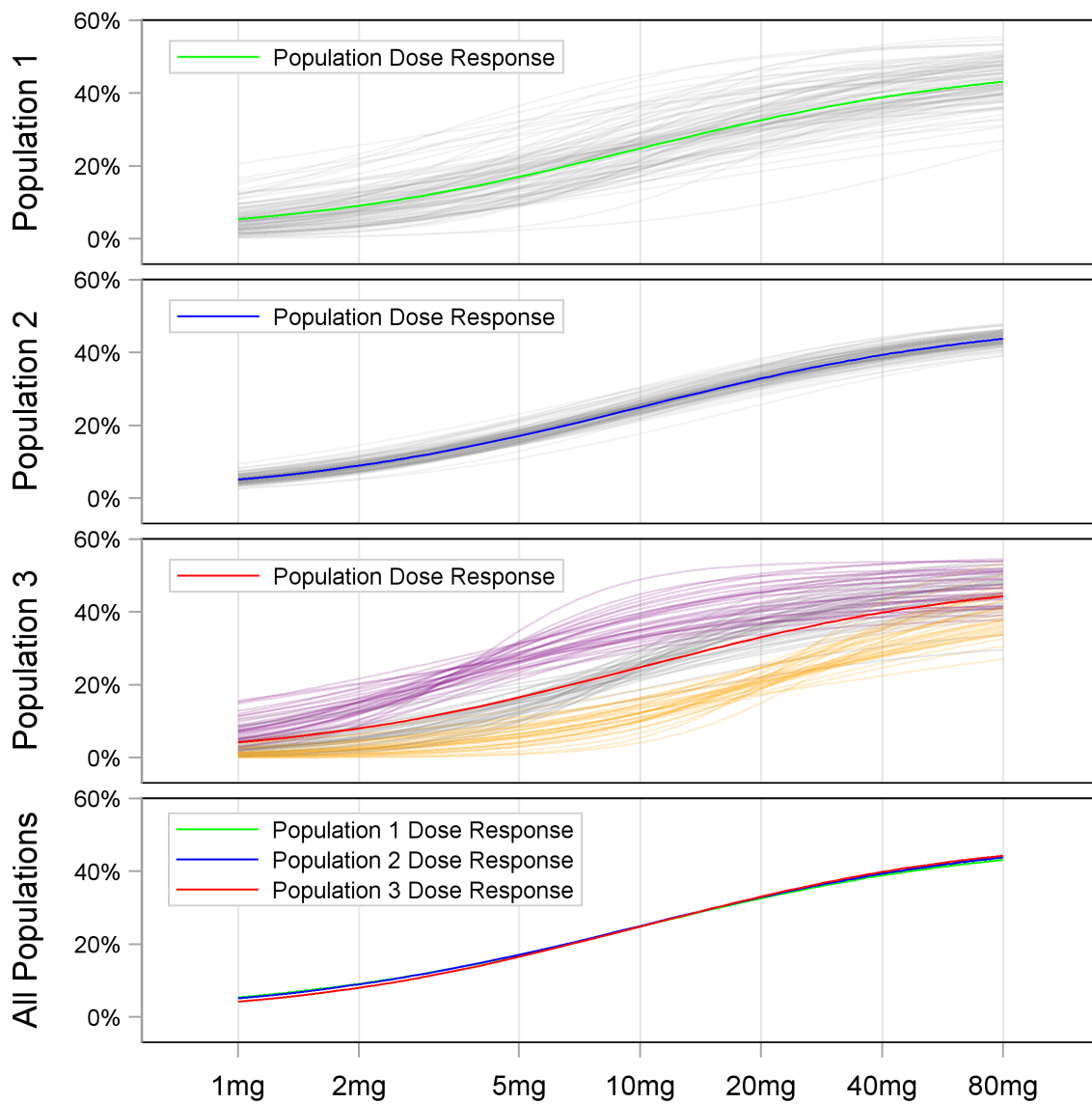
Figure 13.2: Population D-R under 3 different scenarios are shown alongside 100 random individual D-R curves. For population 3, there are 3 subpopulations (colored purple, gray, and orange). The three population D-R curves are shown superimposed in the bottom panel

In this example, as in real-life, attempting to give the same dose to all individuals (say 10 mg) will result in some individuals being under-dosed, and some individuals being over-dosed. In addition, knowing the **Population** D-R does not provide any information to help those patients (like Jill Feldman) who wish to adjust their dose to manage horrendous adverse events; the enormous money spent on such development programs and trials are failing here is answer a very predictable question **"If I reduce my dose due to severe adverse events, how will this change my future outcomes?"** It is perfectly possible that patients who reduce their dose actually **outperform** those that do not, since the former group may have, on average, higher concentrations and be more sensitive to the drug. Thus, a priori, we cannot simply say any downward adjustment of dose in a (non-random) select group of patients will lead to worse average outcomes compared to those patients who remain on the original dose. Our **Population** D-R is telling us nothing useful here, as it does not tell us anything about individual patients (and we care about individual patients, right?).

> **!** Important
>
> Whether we can or cannot measure the **individual** D-R relations, we must **always** acknowledge they exist; patients are **heterogeneous**, and we must therefore **expect** to need to change the dose **and** understand what will happen thereafter.

Quantifying **Individual** D-E-R relationships requires the response be measured for **multiple** levels of dose/exposure from a flexible dosing regimen. With suitable designs and analyses, this will allow quantification of IIV in D-E-R. In some cases, we will be able to accurately adjust the dose to achieve a given response. Examples are doses of anaesthetic drugs being continually adjusted to achieve an appropriate degree of anaesthesia, insulin doses being titrated to achieve appropriate glucose control, and warfarin doses being titrated to achieve an international normalised ratio (INR) within an appropriate range. These examples provide us with a template for drug development, both in terms of providing solutions for difficult problems, and on understanding IIV in response. Firstly, **there is considerable IIV in the doses required to achieve the target responses across individuals (i.e. 10-50 fold dose ranges).** Add references / Show figures? Secondly, it may not always be possible to personalise the dose on the primary responses of interest (e.g. to reduce the risk of micro/macro vascular complications in type 2 diabetes (insulin), or to reduce the risk of serious thrombotic events (warfarin)). Rather, the dose in personalised based on surrogate endpoints/biomarkers known to be correlated with the primary responses (glucose levels in type 2 diabetes, and INR with warfarin). Thirdly, the observed PD responses can take minutes (IV anaesthetic drugs), hours (insulin) or days/weeks (warfarin) before the full cascade from dose to PK to PD are observed, and hence the drug titration strategy needs to simply be structured over a suitable time period based on knowledge of the temporal PK and PD effects. **Fundamentally there is no difference whether the appropriate time interval for dose titration is minutes or months/years.** Finally, in these examples it is common to adjust the starting dose based on individual patient characteristics. For example, the anaesthetist will start with a different initial dosing regimen for a small child compared to a large adult, followed by subsequent dose

titrations. It is worthwhile to consider how these three examples fit the general case shown in Chapter 3 (repeated below).

1) Given the patients individual characteristics, what is the best drug and **initial** dosing regimen?

2) If the initial dosing regimen needs to be changed for efficacy and/or safety/tolerability, how best to do this; what is the **best dose titration algorithm?** That is, based on clinical endpoints, biomarkers, imaging and/or patient reported outcomes (PRO), when should the dose be changed, and by how much.

3) Under what circumstances should the dosing regimen be halted? That is, there is no dose for the patient that has a sufficiently positive benefit-risk to justify continued dosing.

These three examples all show IIV leads to a **range of different optimal doses across individuals**. **There is no "one-size-fits-all" optimal dose**. Here the endpoints/biomarkers of efficacy/safety are precisely measured, and clearly show IIV and the need for personalised dosing. I would encourage anyone working in a different therapeutic area where drugs may only have 1-2 doses approved (e.g. oncology, epilepsy, depression, RA etc.), to ask themselves "Why would I expect **less** variation in the optimal doses for my patients with drug X?". If anything, I think many other indications will have **more** complex cascades from dose to PK to PD across more heterogeneous patients, thus these examples should serve to highlight the omnipresence of IIV in D-E-R for all drugs/indications, and **the crucial need for very wide dose ranges**. Although it may be difficult/impossible to measure IIV in D-E-R for some endpoints, we must nevertheless always acknowledge its presence (and hence influence on the optimal dose for each patient). In cases where surrogate endpoints/biomarkers of efficacy are not yet well determined (e.g. Alzheimer's disease), it may be most reasonably to consider tolerability/safety measures as the primary basis for individualised dose titration.

To complete this section, it is worthwhile to critique two concepts in drug development, that of "the" **therapeutic window** and "the" **maximum tolerated dose (MTD)**. The definitions on Wikipedia are:

> *"The therapeutic window... refers to a range of doses which optimize between efficacy and toxicity, achieving the greatest therapeutic benefit without resulting in unacceptable side-effects or toxicity."*

> *"The maximum tolerated dose (MTD) refers to the highest dose of a radiological or pharmacological treatment that will produce the desired effect without unacceptable toxicity."*

In light of the discussions around **Population** and **Individual** D-E-R, it makes little sense to discuss "**the**" therapeutic window or "**the**" MTD. Both definitions above are deficient/incomplete. Firstly, the definition of "unacceptable" is inherently imprecise (unacceptable to whom? (the patient, the doctor, the regulator, the pharma company)). Secondly, neither specifically define whether this is a **Population** therapeutic window or MTD, or

an **Individual** therapeutic window or MTD. For example, warfarin doses of 2-4 mg may keep the INR for one individual within the (therapeutic window) range of 2-3, but a second individual may need 10-15 mg to keep their INR within the same (therapeutic window) range of 2-3. **Like there is no "one-size-fits-all" dose for warfarin, there is also no "one-size-fits-all" therapeutic window, or "one-size-fits-all" MTD.**

To be precise and meaningful, we must think in terms of the therapeutic window or MTD being **unique** for each individual, and hence we need to add a subscript "i" for each individual "i" (e.g. MTDi). In Figure 13.1, the "therapeutic window$\underline{s}$" would be the doses around the "optimal" doses highlighted for each individual. For 100 individuals, we have 100 therapeutic windows. As a final example, it makes little sense to talk about the therapeutic window or MTD for alcohol. For one individual, the MTDi might be 2 beers, and the MTDi for another individual might be 14 beers. To present "the" (average) MTD as 8 beers is neither helpful nor accurate; it is incorrect for both individuals (...and my MTDi for beer is 5). If we are to personalise dosing for drugs, we are much more interested in determining the distribution of these MTDi's (i.e. think of a histogram showing the MTDi's). That is, showing the lowest MTDi's (individuals who are particularly sensitive to the drug) through to the highest MTDi's (individuals who are particularly insensitive to the drug).

The Population D-E-R relationships shown in Figure 13.2 are often used (incorrectly) to define "the" therapeutic window (e.g. from 5-20 mg) or "the" MTD (e.g. 20 mg, if the safety endpoint response of 5% at 40 mg is considered "too high"). There are two main problems with this naïve approach. Firstly, the 5% response rate at 40 mg **is** an accurate response rate if we give 40 mg (without titration) to all individuals. The individual responses at 40 mg (gray lines) are heterogeneous, including individuals across the spectrum from highly sensitive to highly insensitive to the drug; we are determining the rate assuming we **must** give the same fixed dose to all individuals (i.e. 40 mg). However if the doses were titrated (say from 2 mg => 5 mg => 10 mg => 20 mg => 40 mg) then many (more sensitive) individuals may never reach 40 mg, but rather be adequately treated at lower doses. Hence we should be more interested in the safety response rates at 40 mg **only** for those individuals who are inadequately treated at 20 mg (the least sensitive individuals). Secondly, if we consider a 5% safety risk as too high, we can see the range of doses where the individual responses (the MTDi's) cross the 5% threshold. This is shown as a histogram in the Figure xxx below.

Put in figure

Thus the concept of "**the**" MTD is misguided. In oncology, the use of small 3+3 type trials to determine "**the**" MTD is still ubiquitous, along with the subsequent use of "**the**" MTD in pivotal trials. As a result, many patients with cancer are routinely being under and over dosed based on this outdated and scientifically incorrect notion that all patients are identical. Patients are not identical and there is, and always will be, a range of individual MTDs (the MTDi's). I am puzzled why the concept of "**the**" MTD has persisted for so long. Any basic review of the tolerability and safety data from oncology trials (e.g. like that for trastuzumab deruxtecan in Section 11.1) clearly shows that "the" MTD is clearly neither tolerable nor safe

in a large proportion of patients. "The" MTD does not exist, and the quicker we recognise that, the quicker we can refocus on understanding the MTD for each individual patients (MTDi).

In summary, this chapter has introduced the distinction between **Population** D-E-R relationships that are simply based on "average" outcomes, and explained why these **Population** D-E-R relationships cannot tell us about how individual responses will change with dose. Where possible, we should always aim to understand **Individual** D-E-R relationships, and fully acknowledge that each individual will follow their own D-E-R curve with increasing dose.

---

14 What Should Be The Role For Regulators?

At the end of this chapter, the reader will understand:

- The roles and remit of a **Net Benefit Regulator**

- The roles and remit of a **Scientific Regulator**

- How a **Net Benefit Regulator** only considers average benefits/harms across groups of patients. They do not require any understanding of **Population** and/or **Individual** D-E-R relationships, or whether any fixed or dose titration algorithm is optimal in any way.

- How a **Net Benefit Regulator** implicitly accepts no responsibility for the poor dosing choices (and hence poor patient outcomes) by the pharmaceutical company.

- How a **Scientific Regulator** must determine whether the drug is going to be used in a way that is best for patients. To do this, they seek to understand and precisely quantify **Population** and/or **Individual** D-E-R relationships.

- That a **Scientific Regulator** has a much greater task, responsibility and brings much greater value to both patients and society in general.

- If we care about individual patient outcomes, we need **Scientific Regulators!**

Any discussion on drug development must consider the fundamental question, "What is the role of the regulator?" There are two main options with regards to regulators and the critical roles of dose and individual patient outcomes:

- The **Net Benefit Regulator**: Does the proposed dosing regimen/algorithm yield **Population** benefits that outweigh the **Population** harms?

- The **Scientific Regulator**: Does the proposed dosing regimen/algorithm maximise **Population** and/or **Individual patient** benefits and minimise **Population** and/or **Individual patient** harms?

To be clear, these are two totally different remits; the importance of which cannot be understated.

For the Net Benefit Regulator, the central role of dose is ignored, and individual patient outcomes are ignored. The average efficacy outcomes are simply contrasted with the average safety/tolerability outcomes, with no interest as to whether the dose is 10 mg or 100 mg. The shapes of the dose-response relationships are wholly ignored; the notion of "best use" of the drug is wholly ignored; the importance and influence of IIV is wholly ignored. In contrast, the **Scientific Regulator** has a much greater task, responsibility and brings much greater value to society. **The Scientific Regulator must determine whether the drug is going to be used in a way that is best for patients** or, stately conversely, will **protect** patients from receiving sub-optimal doses, and hence sub-optimal outcomes. Crucially, the evidence base needed for these two roles is completely different.

The **Net Benefit Regulator** may view the world as such. If a drug manufacturer seeks approval for a poorly chosen dose regimen, they may obtain approval based on a modestly positive overall net benefit profile (e.g. think about the many "high dose" oncology drugs that have been approved). In this scenario the subsequent uptake of their drug may be limited, with payers unwilling to reimburse a dosing regimen with mediocre or poor outcomes for many patients. However from the date of approval, patients who **do** receive the drug will be given the poorly chosen dose regimen, and some will suffer the consequences. Over time, physicians/academic groups may slowly learn, perhaps through trial and error, improved dosing strategies that differ from that described in the drug label; the drug label dosing information is then both obsolete and useless. I think such a scenario reflect terribly on both the drug developer and the regulator, because patients will bare the brunt of the inherently "laziness" from all parties. We need sound, science-based justifications for all approved dosing regimens.

As an example, consider a drug regimen that, relative to standard of care, prevents 10 cardiovascular deaths, but with 10 additional intracranial haemorrhages (bleeds within the skull). Based just on a simple assessment of utility (benefits versus harms), the drug could be approved by our **Net Benefit Regulator** (since preventing one death is "worth" having one intracranial haemorrhage). However what is a lower/titrated dose would prevent 9 cardiovascular deaths, but with only 1 additional intracranial haemorrhage? This regimen has a much higher utility, as the benefits more strongly outweigh the harms. In this example, do we want our regulator to simply decide to approve or not approve based on the dose regimen presented to them, or should we expect them to also consider the role of dose (i.e. the utility of different dosing regimens)? Do we not want our regulators to do such quantitative assessments, to ultimately approve the **best** drug regimens that yield the **best** patients outcomes?

The EMA states that the key principle underpinning a medicine's assessment is (with my emphasis in **bold):**

> *"The balance between the benefits and risks of a medicine is the key principle guiding*
> *a medicine's assessment. A medicine can only be authorised if its benefits outweigh*

the risks.

All medicines have benefits as well as risks. When assessing the evidence gathered on a medicine, **EMA determines whether the benefits of the medicine outweigh its risks in the group of patients for whom the medicine is intended.**

While the authorisation of a medicine is based on an overall positive balance between the benefits and risks at population level, each patient is different and before a medicine is used, doctors and their patient should judge whether this is the right treatment option for them based on the information available on the medicine and on the patient's specific situation."

Source: https://www.ema.europa.eu/en/about-us/what-we-do/authorisation-medicines/how-ema-evaluates-medicines

As I understand the existing legislative remits for both the EMA and FDA, their current assessments are as **Net-Benefit Regulators**. As you might guess, I strongly feel this must change. Indeed I suspect most regulators would agree; it must be very disconcerting to be expected to approve any dosing regimen that causing significant and severe adverse events when the overall benefit-risk is modest, and there is no dose justification beyond "this is what we thought looked OK" based on wholly inadequate data (i.e. the sponsor just hasn't bothered to design and run the trials needed to accurately and precisely quantify the D-E-R relationships). Historically there are examples where the FDA has challenged the appropriateness of the proposed dose (e.g. indacaterol), and their newly initiated Project Optimus specifically aims to ensure dose responses are accurately quantified in oncology. **In oncology, it is very common for very high "one-size-fits-all" dose regimens to be presented for approval, even though no sound D-E-R efforts have been undertaken.** Regulators are therefore forced to approve from a "net-benefit" perspective, even though the severity/frequency of a plethora of adverse events could be reduced at both the population and individual patient levels with more appropriate initial dosing and dose titration. The binary decision between approval and non-approval has been rightly questioned by regulators [19], who correctly identify the incoherence between how our knowledge of how best to use a new drug will evolve over time based on accruing information, and their current remit to simply approve or not approve a particular dosing regimen at a single point in time. A more "transitional" pathway would be more aligned with science.

As our understanding of science, pathophysiology and clinical pharmacology has evolved, so do our requirements of our regulators; we need them to be fit for purpose. We need regulators to be laser focussed on dose regimens. Given our modern understanding of heterogeneity between patients in both PK and PD, we need to change the laws that govern our regulators so they are **Scientific Regulators**. This additional remit may require further funding, but is essential if we want dosing regimen algorithms that are patient centric, and deliver the best outcomes for each and every patient.

15 The Two Regulatory Approval Pathways; "Approval P" and "Approval I"

At the end of this chapter, the reader will understand:

- The need for clearly defined regulatory pathways based on Population outcomes (**Approval P**) and/or Individual outcomes (**Approval I**)

- How **Approval P** is based on simple, fixed-dose, parallel group, trial designs that use **Population** (average) outcomes for both benefits and harms. Such trials **are not** aligned with outcome/value based pricing.

- How **Approval I** is based on advanced trial designs where science-based dose titration algorithms are used to optimise the best **Individual** patient outcomes. Such trials **are** aligned with outcome/value based pricing.

- How **Approval I** will lead to Personalised Dosing and better patient outcomes, and can be stimulated if **Approval I** resulted in a 5-year patent extension.

It has been argued that we must look to understand the dose response at both the **Population** and **Individual** levels if we are to truly understand how best to use a drug, and that we need both the sponsor to conduct the right trials/analysis, and a **Scientific Regulator** to work with the sponsor to ensure that the drug dosing regimen algorithm is the best that it can be.

At a recent FDA-ASCO meeting discussing **dose optimisation**, an industry representation cautioned against raising the bar for approval with the requirement to actually understand the D-E-R relationship at approval. Indeed, such an argument can be augmented by stating that the requirement for such information would delay access to badly needed drugs with novel mechanisms of action in areas of high unmet need. Overall, I do not find these arguments fully compelling, but clearly the right trials may require additional time, cost and expertise for both sponsors and regulators alike.

We can hope for change, but I am sceptical that without a clear distinction between drug development approval based on **Population** and **Individual** outcomes, **Drug Development**

For Patients and **Personalised Dosing** will not gain the traction that we need. We need a clear pathway for each; a proposed solution is shown in the Table 15.1.

Table 15.1: Characteristics of pathways to Approval P and Approval I

Criteria	Approval P (Population)	Approval I (Individual)
Trial Designs	Simple	Simple / Advanced
Dosing Regimens Studied	Fixed Doses	Many / Titrated Doses
Patient Population	Narrow or Wide	Wide
Target Outcome Measures	Population/Average Outcomes	Individual Patient Outcomes
Regulatory Review	Net-Benefit	Scientific
Drug Label Dose	Fixed dose	Dose titration algorithm
Personalised Dosing	No	Yes
Fully Aligned with Outcome/Value Based Pricing	No	Yes
Patent Exclusivity	Standard	Standard + 5 years

Most current drug development programs seek **Approval P**, where the focus is typically on simple designs that look at **Population** outcomes in 1-2 key fixed doses, often in a narrow patient population (i.e. with many exclusion criteria used in the randomised clinical trials). Based on efficacy and safety summaries with **Population** outcomes, a **Net-Benefit Regulator** can consider whether the proposed dose regimen(s) is/are, on balance, favourable. Under this pathway, the dosing regimen is neither optimised at either the **Population** or **Individual** level, nor aligned with outcome based pricing. As such, although scientifically limited, it can be reasonable for the regimen to achieve **Approval P**. This would allow early access to the drug, albeit with a dosing regimen that is not optimised in any way.

The table also shows **Approval I**, where the focus has shifted to individual patient outcomes. The trial designs will typically be more complex, and the focus will be on finding the best science-based dose titration algorithm to enable **Personalised Dosing**. The trials may also consider a more diverse/representative patient population (i.e. fewer exclusion criteria), and use more patient specific outcomes (e.g. PROs) that are fully aligned with outcome/value based pricing measures.

Whilst I believe only **Approval I** ensures personalised dosing for both current and future generations of patients, the dual pathways above provide a framework whereby we can transition smoothly from the **Population** system based on fixed doses and **Population** outcomes to the **Individual** system that targets finding the right dose for each patient as efficiently and quickly as possible using an intelligent dosing algorithm.

Approval I should include multiple stakeholders who currently may not interact as fully as we would hope. For example, imagine designing the trial with payers and patient advocacy

groups in the same room as the sponsor and regulators. Could they agree on a trial design that works for all? Fundamentally, everyone wants the same thing; the best outcomes for each patient, thus I am optimist that acceptable solutions could would be found.

To make **Approval I** happen, we need to create a "win-win" environment for all. This is where a patent extension must come in. **If the sponsor does achieve Approval I, a 5-year extension to the patent exclusivity would be granted**. This "reward" for the sponsor would encourage them to invest in the right trials to seek **Approval I**. Their submission fees would help fund additional resources needed by the regulators to assess the scientific integrity of these dose individualisation trials. Both current and future generations of patients (and society in general) would benefit from improved individual outcomes from intelligent, science based, dose titration algorithms, and payers would have the right data to agree reimbursement with sponsors based on value/outcome based pricing.

If **Approval I** were available tomorrow, we would see innovation in exactly the areas we would desire. For example, some sponsors would see the opportunity to extend the revenue cycle of their commercially successful drugs. Although driven my financial motives, the outcome here would still be very positive. These drugs are (demonstrable) good, but we would have the opportunity to gather much better data to truly maximise their benefits via personalised dosing; we would be turning good drugs into great drugs for both current and all future generations of patients (experienced drug developers could point to a wide range of drugs that are frequently prescribed for decades, but with poorly justified dosing regimens with little/no quantitative advice on how to "tailor" the dose based on accruing efficacy/safety data for a patient)).

In areas of high unmet need, such as rare diseases, sponsors would see the importance of **immediately** building in dose modifications for each and every patient, in pursuit of the best outcomes for each and every patient. Initial dose selection, followed by (intelligent) dose titration follows immediately when we seek to maximise the outcome of "this" patient. Conversely, I have seen trial designs for rare indications that are simply fixed-dose phase 3 type trials but in very few patients. These sponsors seem paralysed by the apparent need for a standard **Approval P** program, but the resulting trial data serves no one. Particularly in rare diseases, trying to find the "one-size-fits-all" dose makes absolutely no scientific or ethical sense. The opportunity to seek **Approval I** (and not **Approval P**) would enable these sponsors to fully focus on maximising individual patient outcomes, as both patients and regulators would wish.

It could be argued that this "win-win" approach has a loser - the payer in the extended 5-year period. However I would argue we are (collectively) paying $1 to get, say, $5 in return. The application of **Personalised Dosing** will benefit patients/payers both before and (long) after the 5-year period. Costs, such as hospital admissions and absence from work, will be reduced when we get the dosing right. Patients will be better treated, and be happier. The benefits to society will extend indefinitely.

In summary, the complementary pathways of **Approval P** and **Approval I** precisely clarifies the distinction between population and individual based outcomes, with **Approval I** leading to a modern drug development paradigm that uses **Personalised Dosing** to achieve the best outcomes for each and every patient.

16 The Two Development Strategies; "Strategy P" and "Strategy I"

At the end of this chapter, the reader will understand:

- **Strategy I** uses **Individual** outcomes to obtain **Approval I** for a smart science-based dose titration algorithm.

- That any dose titration algorithm (however complex or simple) to support **Approval I** still produces tabular output for the benefits and harms that are directly comparable to fixed-dose regimens.

- How all stakeholders (regulators, patients, physicians, payers and the drug company) define what endpoints are important, and work with analysts to design the science-based dose titration algorithms.

- How **Strategy P** uses **Population** outcomes to obtain **Approval P** for either:

  - **Fixed-Dose**: Approve 1-2 doses that are "optimal".

  - **Hybrid**: Approve a dose range, with **Personalised Dosing** enabled via a simple dose titration algorithm.

- That **Strategy P** only requires a wide range of fixed-dose regimens to be studied throughout the drug development program.

The previous chapters have made the case for approaching drug development based on a fixed-dose strategy based on **Population** outcomes (herein called **Strategy P**), and/or a **Personalised Dosing** strategy based on **Individual** outcomes (herein called **Strategy I**). Thus **Strategy P** uses **Population** outcomes to obtain **Approval P**, and **Strategy I** uses **Individual** outcomes to obtain **Approval I.**

Now we will discuss, from a non-technical perspective, how both strategies can be implemented, but first we will discuss some of the similarities between the two strategies.

> **! Important**
>
> I think the concept of **Personalised Dosing** based on a dose titration algorithm leads to anxiety/fear from some stakeholders that the data generated will not be easy for them to understand, and hence such drug development strategies should be avoided. **This is misguided.**

In addition, many people in clinical drug development and regulatory groups may only have experience with clinical trials using fixed-dose regimens. For example they may ask, how can one easily see the dose-response, if patients are being titrated to different doses based on measures of efficacy and safety? Indeed, even in a leading journal in clinical pharmacology, Clinical Pharmacology and Therapeutics (CPT), a paper was published in 2021 entitled "The Drug Titration Paradox: Correlation of More Drug With Less Effect in Clinical Data" [20], where the first line of the abstract read:

> *"While analysing clinical data where an anaesthetic was titrated based on an objective measure of drug effect, we observed paradoxically that greater effect was associated with lesser dose"*

Astonishingly, the authors (and the editors!?) considered this a paradox (a logically self-contradictory statement or a statement that runs contrary to one's expectation). Their observation was that, **during the maintenance phase of anaesthesia**, lower doses of propofol were associated with greater effects on the brain (as measured by the Bispectral Index). The simple analysis employed focussed **only** on the **maintenance** dose for each individual, and ignored the within individual titration that would have occurred previously. **It is perfectly normal to expect that response-guided titrations will generally lead to a flat dose-response if only the final dose is used**; it is also not remarkable that a modestly positive or negative dose-response relationship is observed (as observed by these authors), but any such simple analysis is wholly unreliable and meaningless. The publication of this article shows that there remains confusion over dose-response relationships with response-guided dose titration data even within clinical pharmacology circles. Fortunately I recently reviewed another CPT paper that showed an appropriate analysis could avoid such a "paradox!" [15] in case anyone wasn't sure it was possible! Thus for any response-guided dose titration, we cannot just create summary tables by final dose, as these will clearly be wrong, since only some (non-random) sample of patients will be titrated to the higher doses. In contrast, fixed-dose strategies are very easy to understand; unfortunately these are invariably not best for patients, so when we remember that we must put patient outcomes first, we cannot be lazy and, a priori, reject the potentially enormous benefits for patients with individualised dosing. At the highest level, we can think of the results of any individualised dosing strategy at the summary table level, without any regard to the actual dose titration algorithm used. For example, in Figure 16.1 below we combine the general tabular results from a typical drug development program with 4 hypothetical results we might achieve with individualised dosing. The table shows 9 end-points (3 efficacy, 3 tolerability, and 3 safety) and how well or poorly the individualised dosing

algorithm might work relative to the (reference) fixed-dose regimen.

Endpoint	Placebo	Comparator	Fixed Dose	Individualised / Personalised Dosing			
				Result 1	Result 2	Result 3	Result 4
Efficacy 1			Reference	😐😟	😊	😐	😊
Efficacy 2			Reference	😐😟	😊	😐	😊
Efficacy 3			Reference	😐😟	😊	😐	😊
Tolerability 1			Reference	😐😟	😐	😊	😊
Tolerability 2			Reference	😐😟	😐	😊	😊
Tolerability 3			Reference	😐😟	😐	😊	😊
Safety 1			Reference	😐😟	😐	😊	😊
Safety 2			Reference	😐😟	😐	😊	😊
Safety 3			Reference	😐😟	😐	😊	😊
...							
When compared to the reference, 😊 = better, 😐 = similar, 😟 = worse							

Figure 16.1: An illustration of how different Personalised Dosing results could compare to the results from a typical, fixed-dose regimen, program.

Strategy P may look like columns 1-4 in the above.

Strategy I may yield Result 1; dose individualisation has not yielded any improvement relative to the fixed-dose strategy. In this case, either the titration algorithm was poor, or the drug is less amenable to dose individualisation, or both.

Strategy I may yield Result 2; dose individualisation has yielded an improvement in efficacy relative to the fixed-dose strategy, but at no cost in terms of safety/tolerability.

Strategy I may yield Result 3; dose individualisation has yielded an improvement in safety/tolerability relative to the fixed-dose strategy, but at no cost in terms of efficacy.

Strategy I may yield Result 4; dose individualisation has yielded an improvement in both efficacy and safety/tolerability relative to the fixed-dose strategy.

Except for our rare "diamond" drugs like sitagliptin, I would consider Result 1 very unlikely. Result 2 and 3 could come from the same drug, where the dose escalation and dose range was more aggressive/efficacy focussed (Result 2) or more cautious/safety focussed (Result 3). Result 4 is the best outcome, where through dose individualisation we have achieved improvements in both efficacy and safety relative to fixed dosing. **Note Result 4 is what always happens with anaesthetic agents, warfarin, basal insulins, alcohol etc. (I might ask "Why would this not be true for drugs in your therapeutic area?").**

In the above tabular output, the actual dose titration algorithm used is unseen. From a reviewer's perspective, **how** the better outcomes were achieved does not even need to be fully understood. To illustration this point, there are many aspects of drug development I do not

understand, for example the manufacturing processes of monoclonal antibodies. I just know the right people are skilled to ensure their final product, the drug dose of the monoclonal antibody, is correct. Thus when we review the table, we just need to know the right people worked to achieve the best science based dose-titration algorithm. **Thus I hope this may help to allay any anxiety/fear with Personalised Dosing; we just get additional columns with, we may expect, better outcomes for patients (and that is what we really care about!).**

Clearly to effectively design science-based dose titration algorithms, we need multiple stakeholders to work together:

- Regulators, patients, physicians, payers and the drug company to define what PD endpoints for efficacy and safety/tolerability are most important.

- Analysts to propose dose titration algorithms based on measurable endpoints, including surrogate endpoints, biomarkers, imaging data, PROs, drug concentrations and safety/tolerability endpoints. These must be transferable to routine clinical practice.

As an (old!) analyst, I am sure that proposing dose titration algorithms in all therapeutic areas will be possible. In some therapeutic areas it will be very straightforward, with easily measured endpoints being highly predictive/correlated with later outcomes, whilst other areas may require more invasive and/or costly measures (e.g. tumour biopsies, scans) and/or repeated observations due to inherently variable measures. As Lewis Sheiner said "We always know something".

In my world, it would be perfectly judicious to study multiple titration algorithms, for example a "Simple goal" algorithm that is lightweight and easy to use (e.g. using simple endpoints such as patient/physician global assessments after each month). Alternatives could include more complex dose-titration algorithms using multiple assessments (e.g. changes in neutrophils, changes in tumour size etc.), and could be labelled such as "Efficacy goal", where the goal is to focus on getting patients quickly towards their personal MTDi, or "Safety goal", where a less aggressive, more patient friendly, dosing algorithm is used. **Would it not be fantastic to see such trials, allowing us to meaningfully weight up different algorithms in terms of the outcomes they achieved versus their easy of use in clinical practice?** Would this not lead to the best dose titration algorithms for all patients going forward? I think so, and hope you would agree.

To summarise, **Strategy I** will require the development of science-based dose titration algorithms that are focussed on key efficacy and/or safety/tolerability endpoints relevant to the drug and therapeutic area. However the outcomes achieved with these personalised dosing algorithms can be easily compared with those achieved with the simpler fixed-dose regimens. **Drug Development For Patients** needs these trials.

For **Strategy P**, the drug development steps are simpler; we need to study a wide range of doses throughout the drug development program. The following is largely based on the text from a paper I wrote in CPT [21] that introduced **Strategy P**. This was my attempt to

"update" the seminal Learn-Confirm paper by Sheiner [22], where he highlighted the critical role of "learning" in drug development. He commented:

> *". . .the intellectual focus for clinical drug development should be on understanding (i.e., science and learning)."*

In the 25 years since he wrote this paper, the importance of quantifying and predicting the safety/tolerability of different dosing regimens (in addition to efficacy) has become a central component in the evaluation of any new drug. Sheiner did recognize the importance of being able to estimate how safety endpoints change as a function of drug exposure and patient covariates. He wrote:

> *"In confirmatory trials...a larger number of toxicity outcomes may be observed, but this is because the analysis of a confirmatory trial for toxicity is actually a learning analysis".*

The sentiment here is that the "learning" about safety/tolerability is **only** occurring at the end of the confirmative trials for efficacy (i.e. at the end of phase 3), where only 1-2 dose may have been investigated. Today, our trials need to put efficacy and safety/tolerability on an equal footing; we must **plan** to learn about safety/tolerability using our late phase clinical trials in the same way as we **plan** to learn about efficacy. The same logic (i.e., "science and learning") that applied to efficacy endpoints in 1997 must apply equally, if not even more importantly, to safety/tolerability endpoints today. We must therefore design our trials to "learn" how safety/tolerability endpoints change as a function of dose (just like we do with efficacy). D-E-R analyses for safety/tolerability should never be an "afterthought".

When we employ **Strategy P**, the paradigm is one of quantifying D-E-R relationships across both efficacy **and** safety/tolerability endpoints. Thus the design that we want will cover the right dose range and have the best dose levels when we want to accurately and precisely quantify both efficacy **and** safety/tolerability endpoints. This may seem very ambitious, since there are hundreds of potential safety/tolerability endpoints, so how can we design a trial that is "optimal" for all? This is less daunting than it may first appear. The "trick" is to understand that the D-E-R relationships for safety/tolerability will be located either in a similar location to, or to the left of, the D-E-R relationships for efficacy. That is, if the ED50 for efficacy is 10 mg, the ED50 for tolerability/safety endpoint will generally be similar to or lower than 10 mg. In the later chapter on the design of Population D-E-R, we will cover this in much more detail, but the principle here is that we need to think about the precision of the D-E-R for safety/tolerability when designing these trials.

The end result of **Strategy P** will be a set of D-E-R relationships across all efficacy and safety/tolerability endpoints. An example across 3 endpoints is shown below in Figure 16.2.

When reviewing such D-E-R relationships, it is clear we are "trading-off" the increasing benefits with higher doses with the increasing risks with higher doses. As a drug company or regulator, we have the full picture of what our drug is doing, and can therefore make informed decisions
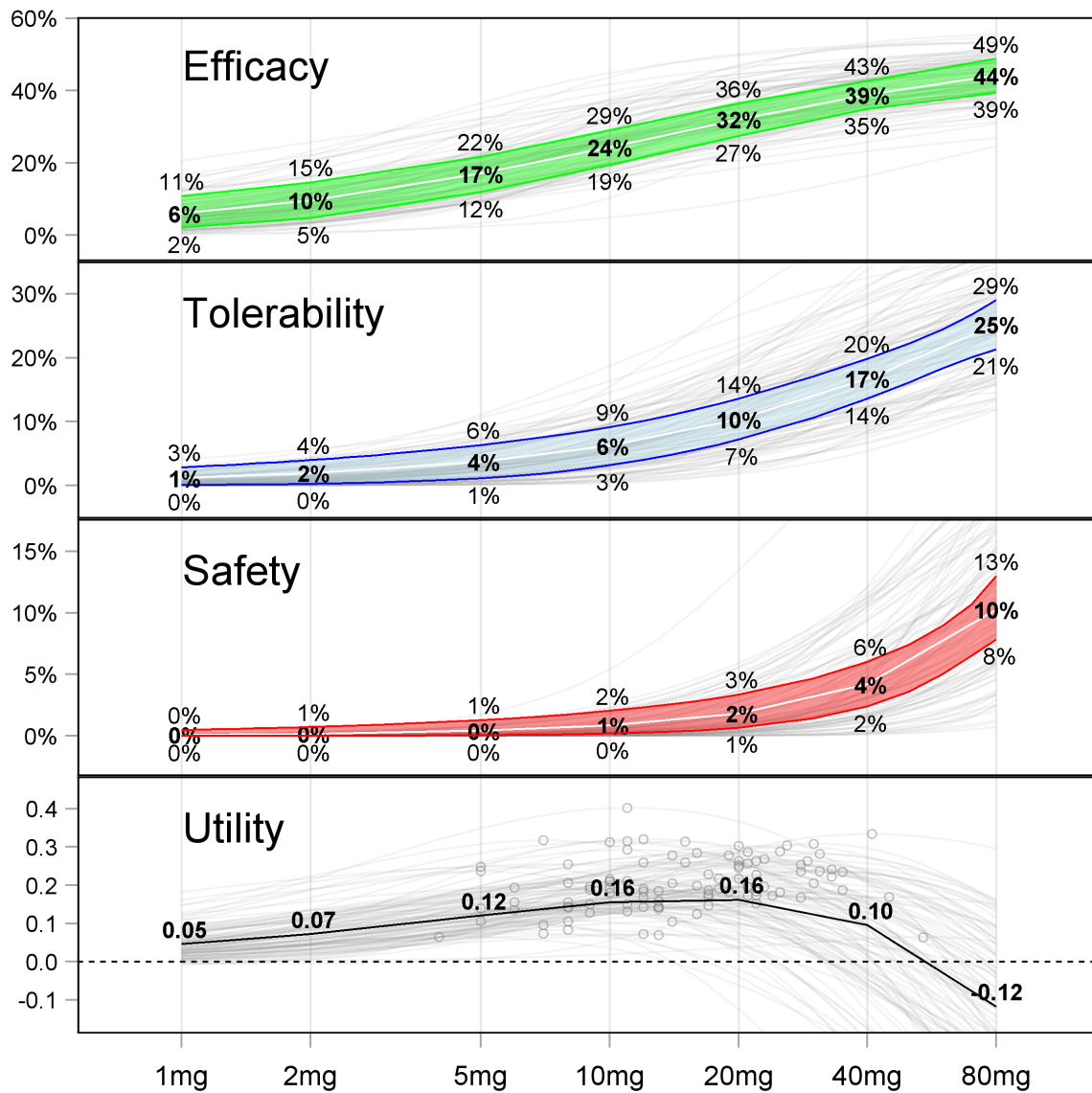
Figure 16.2: The estimated population dose-response (D-R) and 90% prediction interval (white lines and shaded regions) for an efficacy, tolerability, and safety endpoint are shown in the top three panels. The light gray lines are 100 random individual D-R curves; these are typically unobserved but which, when averaged over, generate the population D-R. The fourth panel shows the population utility curve (black line) along with the individual utility curves (gray lines). The circles represent the maxima for each individual utility curve

around what dose(s) to approve. The beauty of the above results is that the drug company has not had to "guess" the magic phase 3 dose(s) based on pitifully small phase 2 trials (recall how most phase 2 trials are not even precisely quantifying the D-E-R for efficacy, let alone for safety/tolerability). Here there is clarity because of the wide dose range and sufficiently large sample size.

Strategy P comes with **two** dose range approval options.

- **Fixed-Dose**: Approve 1-2 doses that (at the population level) are "optimal"

- **Hybrid**: Approve a dose range, with personalised dosing enabled via a simple dose titration algorithm.

Option **Fixed-Dose** is the best "one-size-fits-all" dosing strategy. That is, if we **must** pick just 1 or 2 fixed-dose regimens (with no option to titrate), then these will be our best dose regimens. In the above example, this could be 10 mg and 20 mg, since arguably they represent the fixed-dose regimens with the best benefit-risk profiles.

An alternative to just picking 1-2 fixed-dose regimens is option **Hybrid.** Here we are enabling personalised dosing by supporting a wide dose range, but one based on only the population D-E-R relationships (i.e. a hybrid between individual and population strategies). In the above example, this could be to approve the range 2.5 mg, 5 mg, 10 mg, 20 mg and 40 mg. Here we are implicitly recognising that the "one-size-fits-all" doses are going to be too high or too low for some patients, because we understand that there are individual D-E-R relationships "behind" these population D-E-R relationships (the gray lines in the figures). For example, in RA or epilepsy, the patient could start on the 2.5 mg for the first month, before moving to 5 mg / 10 mg / 20 mg / 40 mg in months 2 / 3 / 4 / 5 respectively if required.

> **!** Important
>
> Compared to **Fixed-Dose**, the **Hybrid** approach is most consistency with the "**do no harm**" principle.

If we push all patients directly onto 10 mg from day 1, there will be patients who are particular sensitivity to the drug (low individual ED50s for efficacy and tolerability/safety) and/or with higher than average drug concentrations relative to the average patient (low individual clearances). For example, one of these patients may experience a moderate or severe adverse event at 10 mg. If this patient had started at 2.5 mg, the intensity/severity of the adverse event would be much lower (or they may not experience the adverse event at all). Indeed, this patient may be perfectly well treated at 2.5 mg and hence not need to titrate to a higher dose. Thus **Hybrid** will reduce the incidence, frequency and severity of adverse events, and allow patients to find "their" dose across the dose range. The only limitation of **Hybrid** is that it will take a little longer for the least sensitive patients to find their dose as they titrated from the lower doses to the higher doses.

The **Hybrid** option may initially seem more onerous, requiring physicians and patients to meet regularly after each dose level to decide whether to increase/maintain/decrease the patients dose level. However this could be achieved quite simply. For example, under the **Fixed-Dose** option, the patient may be sent home with a box containing a 3-month supply of 10 mg, will the advice to call if they experience any adverse events during this time. Under the **Hybrid** option, they could be sent home with a box containing the month 1 supply of 2.5 mg, the month 2 supply of 5 mg and the month 3 supply of 10 mg, will the same advice to call if they experience any adverse events during this time. Unless the 10 mg dose has an excellent safety profile (like sitagliptin 100 mg) **I see the latter as most coherent with the important "do no harm" mantra**. Most drugs do not have an excellent safety profile, and hence why the **Hybrid** approach is appealing.

In the discussion above, the simple titration algorithm was to increase the dose at monthly intervals. Clearly the timeframe for the dose titrations under the **Hybrid** option would be based on knowledge of the magnitude and temporal changes in both the PK and PD effects, and hence could be short or longer than at monthly intervals. This would naturally be balanced against the need to find the right dose for each patient as quickly as possible.

This chapter has outlined the following approaches to drug development and approval:

Strategy I uses **Individual** outcomes to obtain **Approval I** for the optimal science-based dose titration algorithm.

Strategy P uses **Population** outcomes to obtain **Approval P** for either:

- **Fixed-Dose**: Approve 1-2 doses that are "optimal".
- **Hybrid**: Approve a dose range, with personalised dosing enabled via a simple dose titration algorithm.

From a trial design perspective, both **Strategy P** approaches require the same trial designs, thus their difference is in how all stakeholders (regulators, patient/physicians, drug companies etc.) decide whether the **Fixed-Dose** or **Hybrid** will ultimately serve patients/physicians best. Although the drug company may see the need to support multiple dose strengths as an additional burden, I would encourage them to see this as a small price to pay to enable each patient to find the right dose for them, and hence happily stay (and pay) for the drug going forward (recall lower "churn" = greater revenue).

In later chapters we will discuss in more technical details about the design and analysis of the D-E-R trials needed to support both **Strategy P** and **Strategy I.**

17 Changing How We Pay For Drugs; Value/Outcome-Based Pricing And Subscription-Based Pricing

At the end of this chapter, the reader will understand:

- That expecting reimbursement/payment for a drug for a **particular** patient based on "average" outcomes in **other** patients is odd.

- What is value/outcome-based pricing, and how it should refocus drug manufacturers to use personalised dosing to achieve the best outcome for each and every patient.

- It is in the drug manufacturers commercial interests to produce a wide dose range to enable each patient to find the dose that works best for them; otherwise they lose the patient, and further revenues, forever.

- For some indications/drugs, it would be more sensible to employ a subscription-based pricing, whereby the drug manufacturer is paid whilst the patient is (happily) using the drug (i.e. the subscription price is independent of the dose/usage).

When it comes to how we pay for drugs, consider this analogy. Imagine that you buy a new TV. When you plug in the TV, you get an awful picture. You call the manufacturer, and they take your personal details. They respond "Well, that is interesting. According to our data, we have many people like you who did achieve an acceptable picture". This would be a comical interaction; as a consumer, I have no interest in whether a product worked for "people like me" or works "on average"; I would be very unhappy with the product, and demand my money back. I hope this analogy brings into sharp focus the inherent shoddiness in how drug manufacturers often expect to be reimbursed/paid for their products. I cannot think of another industry that expects full payment when their product fails to deliver. The introduction of **value/outcome-based pricing** will rightly bring much greater scrutiny to individual patient outcomes.

> **!** Important
>
> **Value/outcome based pricing is where the reimbursement for the drug is directly linked to the value/outcome achieved for the patient.**

In the most basic sense, if the drug fails to deliver real benefits for the patient, the manufacturer does not get paid. An excellent and fair system!

I view the emergence of value/outcome-based pricing as a "game changer" in how the pharmaceutical industry must approach dosing, and the central role of personalised dosing. The industry has a choice. Either develop a "one-size-fits-all" dose based on "average outcomes" that is poorly aligned with value/outcome-based pricing, or develop a smart dose-titration algorithm (with a wide range of potential doses available to each patient) aimed at maximising individual patient outcomes that is fully aligned with value/outcome-based pricing.

add a graph showing a range of outcomes based on fixed-dose regimen, and the higher range with individualised?

From a purely commercial perspective, it is clear that offering multiple dose levels will enable multiple opportunities to achieve success for the patient, and hence the greatest chance of continued reimbursement.

Having multiple doses will require additional manufacturing requirements and devices such as the injection pens used so successfully with insulin glargine, but this "cost" will be minor relative to the overall cost of the drug, and the value that it brings to both the patient and the manufacturer (longer patient usage). In short, companies need to prospectively develop and utilise commercial products/devices that support a very wide range of dose levels (=support personalised dosing).

Unfortunately not every therapeutic area will have a clear, objective measure of value. For example, whilst a type 2 diabetes drug could be reimbursed relative to the magnitude of the improvement in glucose control (e.g. the % reduction in HbA1c), for an oncology drug the primary endpoint could be the overall survival of the patient. Although alternative measures of value could exist (e.g. reduction in tumour size), we can consider alternative reimburse models that fully enable personalised dosing.

> **!** Important
>
> **We can reimburse drugs based on a subscription-based pricing model.**

The streaming services provided by Netflix and Spotify require the user to pay a fixed price for access to the service, and therefore the price paid is independent of usage. Similarly my daughter pays a monthly fee to the opticians for monthly contact lenses and yearly check-ups. It is in the interest of my daughter, the optician and the manufacturer to ensure her needs are continually met every year. In particular, if her prescription changing over time, the

cost remains the same (the manufacturer has developed/planned to support a wide range of prescription strengths). For some drugs/indications, we could apply a similar **subscription-based** pricing model. The reimbursement would not be linked to the actual dose, but rather on having access to the drug (i.e. the price paid is independent of the dose). Whilst the patient uses the drug, the company is paid. The beauty of these relationships is that all parties benefit when the patient receives the best dose, and hence outcomes, for them.

---

18 What Product Should Drug Companies Sell? (can be skipped)

At the end of this chapter, the reader will understand:

- How historically drug manufacturers final product is simply a dosing regimen and a drug label; these are insufficient if we want to best use the drug for each patient, and very poor value relative to the cost of drug development.

- How drug approval could use modern technologies to bring much more relevant information to the physician/patient that goes far beyond that offered by current drug labels.

- At approval, a program/website with an interactive/smart, science-based, dose titration algorithm could accompany the drug label, helping physicians deliver personalised dosing.

- Patient engagement, adherence and PROs data could, with permission, be collected across willing patients to enable multiple stakeholders (regulators, patient advocacy groups, payers) to further learn how different dosing regimens perform via Real World Evidence (RWE).

For the reader short on time, this chapter can be skipped; it may not be so useful. The aim of this chapter is to elaborate on what should/could be the final "product" of drug development. When I use my Apple computer, it has both hardware and software components. The computer user wants the best outcomes from their software, and without great software the wonders of the amazing hardware cannot be realised. If drug companies did computers, I would say they spend >99% of their R&D dollars on the hardware (making the drug and running the trials), and <1% on the software (providing the user with the right tools to unlock the full potential of the hardware). It is analogous giving someone the latest iPad with only Solitaire installed.

Historically the product the drug company sells is a dose of the drug (often using a price-per-dose (PPD) model). Despite the enormous R&D costs associated with developing a drug, the final product only comes with a drug label, a document that provides limited value to both the physician and patient.

The interaction/feedback between the user (the patient/physician team) and manufacturer (the drug company) is quite non-existent. How is the UX (user experience) being used to improve the product? Do drug companies ever hire UX people to improve their product and customer experience?

What if the product was more than simply a dose? The application of a science-based dose titration algorithm/program is one obvious extension, allowing the patient/physician to tailor the dose over time using, for example, a web-based tool (e.g. hosted by the FDA/EMA). However why not be more aspirational? Could we not aim for integrated/smart devices that could work with an app to record the time and dose of administrations and/or record PROs (e.g. the incidence and severity of key safety/tolerability measures)? These could prove immensely valuable both in terms of adherence and in allowing patients/physicians to accurately monitor the utility of the current dosing regimen. With patient consent, this data could even be shared (anonymously) with other patients/regulators/manufacturers and aggregated/analysed to provide real world evidence (RWE) of usage, and hence provide greater insight beyond the original clinical trials (e.g. imagine an app for patients with lung cancer that allows them to "chart" their tolerability profile and compare it to others, and see how these profiles changed for patients that titrated their dose).

We should look to learn from companies like Tesla and Apple to see how they collate user data to continually improve their products for current and new customers, and devices like the Freestyle Libre 2 System that continually records blood glucose levels in type 1 diabetes to provide forward looking predictions on glucose levels. People can be highly engaged to record/track their personal health data over time, whether that be via a running app, or an app designed for their particular illness, but we need to enable this. Finally, patient advocacy groups/charities/non-profits representing patients would truly appreciate these interactive apps/tools/dashboards, with potentially high engagement from patients to "make things better" and learn for both themselves and future patients.

As an aside, I would like to make some observations on modern drug labels. The goal of the label is to highlight important information to the "end user", the prescriber/patient. However the sequential presentation of copious pieces of information actually detracts from the overall usefulness of the prescriber/patient; they "cannot see the wood for the trees". I see an analogy with my consulting work. On the one hand, I would like to include in my presentations sufficient details and results to ensure the full analysis and results are clear. However in doing so my "product", the slide deck, can become far too large; the "end user" (the project team) do not have 3 hours to go through everything, but rather just need the key results (thus my additional material goes into the backup). Similarly, if we want the drug label to be most useful to prescribing physicians, are they really the best they could be? For example, the current FDA label for pembrolizumab (brand name Keytruda) extends to **106** pages; it provides copious information on a wide range of topics: the recommended dosing information for over 15 types of cancer, severe and fatal immune-mediated adverse reactions and laboratory abnormalities, when treatment should be withheld or discontinued, various drug combination

and interaction information, clinical trial results etc. To illustrate how verbose and unwieldy the label is in parts, consider (a part of) the dosing information shown in Figure 18.1 below.

There is a **lot** of repetition in the above.

Imagine a physician is considering prescribing pembrolizumab in combination with carboplatin to a women of child bearing potential for the treatment of metastatic squamous Non-Small Cell Lung Cancer (NSCLC). They are specifically interested in the appropriateness/justification of the 400 mg every 6 weeks regimen (pages 54 and 99), concerned about fetal toxicity (page 52), and the dosing regimen used in the corresponding clinical trials (page 61). In a world where we are all rather impatient, this is painfully tiresome; trying to navigate the verbose drug label to find this information is both laborious and prone to error.

Now imagine that the same physician goes to the new FDA website for approved drugs, including pembrolizumab. Here they can directly select the drug and type of cancer, immediately obtaining **only** relevant information (think how 80% of the label has now been "moved to backup"). Now consider additional input fields that can capture key information relevant to their patient (sex, weight, previous treatments, renal function etc.) that can further be used to highlight **relevant** key results (e.g. fetal toxicity for women of child bearing potential). In addition, there could be "tabs" allowing the selection of real (anonymised) data from similar patients for key safety endpoints (e.g. rather than simply warn of the risk of neutropenia, show actual trial results of neutrophil counts over time for, say, women 25-35 years of age (like the patient)), with dynamic "dashboards" showing the risk of each grade of neutropenia as a function of different dosing regimens (numbers that are updated when combined with neutrophil counts from the patient after the initial treatment cycles etc.).

At the start of this book I discussed what would drug development look like if we could redesign it from scratch. I feel similarly about drug labels; I understand the history and understand the current format. However I am convinced we can do so much better. Primarily, we must put the needs of the end user (prescriber/patient) at the centre of everything we do. What do they need to know to ensure the drug dose regimens are used as effectively as possible? How can we help them, whilst also ensuring key safety information (e.g. contraindications, drug-drug interactions etc.) is also highlighted? How can we use technology to present the material as cleanly and crispy as possible? I think if all stakeholders and technology experts could sit down and prototype an interactive, web-based drug label, I am sure the final product would be light years ahead of, for example, the 106 page behemoth that is the current pembrolizumab label. We need more UX!

In summary, when we think about the final product of drug development, we should broaden our vision beyond simply a drug dose and drug label. The integration of modern technologies will substantially and continually improve the value and performance of the "product" to patients and provide informative and quantitative information to the prescriber. These tools combined will ultimately leading to better patient outcomes.

-------------------- DOSAGE AND ADMINISTRATION --------------------

- Melanoma: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- NSCLC: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- SCLC: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- HNSCC: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- cHL or PMBCL: 200 mg every 3 weeks or 400 mg every 6 weeks for adults; 2 mg/kg (up to 200 mg) every 3 weeks for pediatrics. (2.2)
- Urothelial Carcinoma: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- MSI-H or dMMR Cancer: 200 mg every 3 weeks or 400 mg every 6 weeks for adults; 2 mg/kg (up to 200 mg) every 3 weeks for pediatrics. (2.2)
- MSI-H or dMMR CRC: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- Gastric Cancer: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- Esophageal Cancer: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- Cervical Cancer: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- HCC: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)
- MCC: 200 mg every 3 weeks or 400 mg every 6 weeks for adults; 2 mg/kg (up to 200 mg) every 3 weeks for pediatrics. (2.2)
- RCC: 200 mg every 3 weeks or 400 mg every 6 weeks with axitinib 5 mg orally twice daily. (2.2)
- Endometrial Carcinoma: 200 mg every 3 weeks or 400 mg every 6 weeks with lenvatinib 20 mg orally once daily for tumors that are not MSI-H or dMMR. (2.2)
- TMB-H Cancer: 200 mg every 3 weeks or 400 mg every 6 weeks for adults; 2 mg/kg (up to 200 mg) every 3 weeks for pediatrics. (2.2)
- cSCC: 200 mg every 3 weeks or 400 mg every 6 weeks. (2.2)

Figure 18.1: The Dosage and Administration section of the FDA label for Keytruda curve

19 Adaptive Randomisation In Population D-E-R Trials; Why We Should Learn As We Go

At the end of this chapter, the reader will understand:

- To accurately and precisely quantity the Population D-E-R relationships as efficiently and quickly as possible, doses should be placed at the optimal dose levels (the doses that are most informative).

- These optimal dose levels depend on the true Population D-E-R relationships that we seek to quantify; thus to be most efficient (and ethical) we must "learn as we go".

- **Adaptive randomisation** is the most efficient way to "learn as we go". We start with a randomisation schedule that supports a very wide dose range, and then "zero in" on the optimal (most informative) doses.

- To implement **adaptive randomisation**, we need a team similar to a Data Monitoring Committee to review the accruing data and update the randomisation allocations.

- **Adaptive randomisation** is most valuable when our initial assumptions about the location and shape of the D-E-R relationships are imperfect (as they always are!).

Unsurprisingly, in drug development we need to investigate the right dose range for each drug. In addition, selecting the optimal dose levels within this dose range requires an understanding of the location and shape of the D-E-R relationships. We appear to be a chicken and egg situation; **we can only optimally design a D-E-R trial with the information that we are seeking to obtain from the trial!** The pharmaceutical industry is not the first industry to face such a challenge. For example, when Amazon introduces 100 new products, they have very limited understanding of how popular each will be, and hence how many units of each product they should order/stock in their warehouses, and ultimately how much profit each product will generate over the first year. What Amazon **does not do** is wait until the end of

the year to assess the sales of each product. Such a foolish strategy would see them holding far too much stock of the least popular products, and failing to order additional stock for the most popular products. They would not be happy to see the most popular new product sell out by March, but delay reordering until December. **The solution to this problem is to look at the accruing data on a continuous basis, and act accordingly**. Thus Amazon may start with 100 new products, but will use daily/weekly sales to increase their orders for the most popular/profitable products, and reduce the orders for the least popular/profitable products. For the weakest products, these would be phased out of stock, and replaced with the more successful products. This may all seem quite obvious, and it is. Why would Amazon be so stupid as to wait until the end of the year to review their own sales data and act accordingly? Such a lazy strategy would be sure to be less successful than one that does adapt their product lines and orders.

Now if we substitute dose range for products, dose levels for the most popular/profitable products, and the warehouse for patients, I hope the analogy with Population D-E-R trials becomes evident. Initially we start with a wide range of doses (products). As we obtain the accruing data across this wide range, we can "zero in" on the best doses (most popular products), with new patients (stock) being randomised to (ordered for) the best doses (products).

> **!** Important
>
> The above process of "learning as we go" is describing one type of **adaptive randomisation**. It is the most efficient (and hence ethical) way to learn about Population D-E-R relationships.

Thus in a **Population** D-E-R trial we may start with 10 dose levels (e.g. placebo, 0.5mg, 1 mg, 2.5 mg, 5 mg, 10 mg, 20 mg, 50 mg, 100 mg and 200 mg), but ultimately end with a range that is more refined to the actual location of the D-E-R relationships, which may be towards the lower 5 doses (e.g. placebo, 0.5 mg, 1 mg, 2.5 mg, 5 mg and 10 mg), or towards the higher 5 doses (e.g. placebo, 10 mg, 20 mg, 50 mg, 100 mg and 200 mg). In these two simple cases, the two final dose ranges investigated are 20-fold different (i.e. 0.5-10 mg and 10-200 mg). There is a wide "margin of error" when we use **adaptive randomisation**; we are maximising our chances that we do indeed investigate the right dose range/levels.

Unfortunately this is not what happens in >95% of D-E-R trials run by the pharmaceutical industry. Instead, we wait until the end of the trial (=end of the year for Amazon) and then look at the data, to see if our initial guesses were reasonable. **This is painfully wasteful in time, money and patient resources**. To be direct, it is both inefficient and unethical in equal measure. Indeed, I am sceptical of how many D-E-R trials actually prospectively define their "initial guess" of the true location and shape of the D-E-R relationships for efficacy and safety, and then evaluate whether their trial design (dose levels and N) is actually capable of accurately and precisely quantifying the D-E-R relationships. Rather, most D-E-R designs I see appear to select a few convenient doses (often very closely spaced) and small N, and "see what they will get". The interpretation of the final observed data is predictably unpredictable.

With a poor design, it can be easily shown using simulation/re-estimation methods that the observed outcomes in the trial could support a wide range of "best" doses based on simple/naïve analysis (e.g. just picking the dose that happens to have the "best" observed outcomes). As well as wasting both the time and money of the sponsor by running such a trial that yields so little useful data, we must question the ethics of such trials.

When a patient enters such a trial, I think there is an onus on the sponsor to ensure the data generated from that patient will meaningfully contribute to the subsequent analysis. For example, is it ethically acceptable to randomise a patient to 0.5 mg if data that has already been collected on other patients across a wide dose range would show that 0.5 mg is essentially uninformative to our D-E-R understanding? It is my view that we should feel compelled to ensure the data from each and every patient is as informative as it can be; this means ensuring patients are receiving the most informative doses. In a mathematical sense, we can actually quantify how informative each potential dose level is, so for example 5 patients at 0.5 mg could be the same as 1 patient at 10 mg (as data in the "wrong" part of the D-E-R curve is much less useful than data in the "right" part of the D-E-R curve). Thus randomising the patient to 0.5 mg is both wasteful and, in my mind, unethical.

> **!** Important
>
> Adaptive randomisation can be motivated on purely economic grounds, where we **learn the true location and shape of the (true) D-E-R relationships** as **quickly** and **efficiently** as possible (saving both time and money), but also from the **ethical** perspective, where each patient is contributing as much as possible to the goal of the trial.

Two key points of **adaptive randomisation** need to be clarified:

- What is meant by the best/most informative doses?

- Who is unblinded to what data, and how are they changing the randomisation scheme?

When we discuss the best or most informative dose, we need to be precise in what this actually means with respect to D-E-R trials. When our goal is to accurately and precisely determine the D-E-R relationships, we need data at key doses on the D-E-R curve. For example, since the D-E-R for efficacy quantifies the magnitude of the changes versus placebo, we will indeed need some data at the bottom of the D-E-R curves (i.e. low doses and placebo). Thus by best, we mean most informative. We do not mean anything to do with "best" for a particular patient. Remember, these trials, like most clinical trials, are not explicitly trying to give the patient the "best" dose. This is not to say that the dose the patient is assigned to, or in the case of dose titration trials the dose they are titrated to, will not be "best" for them, but this is not our primary goal **during this stage of clinical drug development;** we are still in the learning phase. Medical ethicists have written much on the important interplay between the social and clinical value of clinical trials and the benefits and harms to individual trial participants. I plan to add a bonus chapter on this key topic to share with you my opinions; however in

short, I subscribe to the viewpoint that we must look to minimise any harms to individual patients, but otherwise their participation in D-E-R trials is to facilitate the reduction on the uncertainty around the D-E-R relationships as effectively as possible. **Thus I view each trial participant (patient) as a kind person who may derive no, some, or major benefit from the dosing regimen given to them, but who unquestionably will contribute to our better understanding of how to best use the drug going forward.**

If we wish to use the accruing data to guide the adaptive randomisation, we will require unblinded experts to both analyse and interpret the accruing data, and to update the randomisation allocations. The team of experts will act in a similar role to an (independent) Data Monitoring Committee (DMC) (also sometimes referred to as a Data Safety Monitoring Boards (DSMB)). Regulatory advice on DMCs has been developed (see link below).

[https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-data-monitoring-committees_en.pdf]

To be clear, it is in interest of all parties that we "learn as we go"; the drug company, the regulator, and the patients all benefit with this intelligent approach to running the trial. The drug company are most efficiently using their R&D dollars to quickly, accurately and precisely quantify the D-E-R relationships that both they and the regulators want to know. Patients within the trial are contributing fully to this effort and they, and future patients, may benefit from the best dosing regimens being approved based on a clear quantification of the D-E-R relationships for both benefits and harms for the drug.

The primary tasks of the team would be to:

1) Adapt the randomisation schedule based on the accruing data

2) Stop the trial for futility and/or efficacy

The ASTIN trial [23] **is an excellent example of this type of adaptive D-R trial.** In ASTIN Pfizer sought to determine the D-R of UK-279276 (a neutrophil inhibitory factor) on the Scandinavian Stroke Scale in patients with an acute ischemic stroke. They considered placebo and 15 dose levels from 10-120 mg. Although the D-R model was not something I would endorse, **the adaptive design, conduct and ongoing analyses they employed were all excellent.** Unfortunately the trial ended early for futility, as although the higher doses were targeted more heavily as the accruing data suggested poor efficacy across the dose range, ultimately all doses were insufficiently different from placebo to merit further investigation. **From an operational perspective, this trial was remarkably successful**; Pfizer learnt quickly and efficiently that the D-R for this drug was very poor across the whole dose range. I understand that senior management in Pfizer conflated the performance of the drug (poor) with the performance of the trial (excellent), ultimately deeming this a "failure". Clearly paying for this trial but not getting an approval at the end would have been disappointing, but "failing fast" is a mantra of any business that develops high risk/high reward products, and this trial did its job brilliantly, and should be applauded. Had the drug worked, I think

we would have seen an enormous uptake in these designs, with Pfizer leading the way (a real lost opportunity in my opinion).

I co-wrote a paper showing how **adaptive randomisation** can be used in combination with optimal design for D-R modelling [24]. The work showed that when our initial assumption of the location (ED50 = 10 mg) and shape (Hill=1) of the D-R was exactly correct, the adaptive design will, as expected, stick to the initial (optimal) dose levels that will most precisely estimate the D-R relationship. However in scenarios when the true location of the D-R was different to our initial assumption (e.g. ED50 = 5 mg or 20 mg), or the shape was different (Hill=0.5 or 2), **adaptive randomisation** would learn from the accruing data and adjust the dose levels accordingly, ultimately finding the best/most informative dose levels under the true D-R relationship.

The bottom line here is that employing **adaptive randomisation** to "learn as we go" is always the most efficient strategy, and is extremely valuable when our initial expectations on the location and shape of the D-R are incorrect. Given that is normal **not** to have an excellent understanding of the true location and shape of the D-R prior to running a D-R trial, I am astonished such trials are not routinely used.

_____

20 The Half-Time Summary; What Have We Learnt, And What Solutions Are Outstanding?

In the introduction, I introduced 5 key themes to enable a **Drug Development for Patients** approach to drug development**.**

1) Understanding the goal

2) Understanding the science

3) Understanding dose-response models/modelling

4) Changes needed within the regulatory agencies

5) Changes needed within the pharmaceutical industry

Firstly, I covered why Patients Outcomes Must Come First. Here I motivated the need for personalised dosing, and explained what this meant. When we put individual patients outcomes first, the 3 steps we seek to understand in drug development are:

1) Given a patient's individual characteristics, what is the best **initial** drug and dosing regimen?

2) If/when the initial dosing regimen needs to be changed for efficacy and/or safety/tolerability, how best to do this; what is the **best science-based dose titration algorithm?** That is, based on clinical endpoints, biomarkers, imaging and/or patient reported outcomes (PROs), when should the dose be changed, and by how much.

3) Under what circumstances should the dosing regimen be halted?

This set out our goal – getting the best outcomes for each and every patient by adapting (personalising) their dose.

I then gave a short review of the history of the design and analysis of RCTs. This covered what was good (e.g. randomisation, blinding, control groups), but then explained why we need to move beyond "agricultural" experimental design based on average patient outcomes to one where we refocus RCTs on individual patient outcomes; remember, patients are not fields! A brief history and overview of current "dose-ranging" trials was discussed, explaining how

126

ICH E4 has wrongly led to simple, fixed-dose designs being favoured over more informative dose individualisation designs. Some modern D-R trials were reviewed, and shown to be very weak in terms of their design and analysis for both efficacy and safety/tolerability. These chapters served to provide the necessary context to explain the interrelationship between three choices/directions for drug development:

1) Individual patient outcomes versus Population (average) patient outcomes

2) Individual D-E-R relationships versus Population D-E-R relationships

3) The right designs and analyses of RCT to investigate the above and support drug approval.

Historically we have relied on Population (average) patient outcomes from fixed-dose regimens because such a strategy is very simple. Unfortunately such a basic strategy does not achieve our goal; we do not get the best outcomes for individual patients.

The science of patient heterogeneity was then covered. This explained how IIV in PK and PD are the drivers of **why** patients need different doses, and how dose is just a very crude mechanism by which we seek to deliver sufficient drug to the site of action to illicit the desired PD responses in a patient. From anaesthetic agents to alcohol, we are all aware how different people need different "doses"; for most drugs, only dose individualised can deliver the best outcomes for each and every patient. We also viewed personalised dosing more from the patients perspective, and how they should be able to make decisions around their dose with consideration to their treatment goals, outcomes and personal preferences. As drug developers/regulators, we should ensure all approved dose ranges are scientifically justified, but we cannot tell people how they should feel.

To understand D-E-R relationships, we must use D-E-R models. The effects at different doses are **always** related to each, and suitable D-E-R models, such as the excellent sigmoidal Emax model, provide the best mechanism to allow us to "link" the effects across the different doses. Since the shape of D-E-R relationships are often poorly understood, I explained what "flat", "typical" and "steep" D-E-R relationships look like for most clinical endpoints we encounter in drug development. Integrated D-E-R modelling across all doses/data is right, whilst "cherry picking" the observed outcomes at individual dose levels was shown to be unscientific and wrong.

The crucial differences between Population (average) and Individual D-E-R relationships were investigated further, explaining how the quantification of the Population D-E-R does not tell us about the shape of Individual D-E-R relationships. Where possible, we should always aim to understand individual D-E-R relationships, and fully acknowledge that each individual will follow their own D-E-R curve with increasing dose. **Individual** optimal doses are markedly different to any **Population** optimal dose. Like there is no single "optimal" dose across all patients, there is also no single therapeutic window and no single MTD. Many patients suffer terribly at "the" Population MTD, because this dose is much higher than their own MTD (MTDi).

In summary, the science tells us we need to consider **Personalised Dosing** for each patient. Unfortunately the same simple fixed-dose regimens for all patients will inevitably lead to poor outcomes for many patients; we can and must do better.

As our understanding of science, pathophysiology and clinical pharmacology has evolved, so must the remit and responsibilities of our regulators. To serve patients and society best we need, and must fund, **Scientific Regulators**. I explained how a **Scientific Regulator** must determine whether the drug is going to be used in a way that is best for patients; they will demand to see the accurate and precise quantification of Population and/or Individual D-E-R relationships for efficacy, tolerability and safety endpoints. A **Scientific Regulator** will advocate that such evidence is infinitely superior to simple "by trial, by dose" tables, listings, and P values. We cannot continue to accept **Net Benefit** regulators being compelled to approve drug regimens based on a (marginally) favourable benefit-risk assessment when there is little or no justification for the proposed dosing regimen; poor dosing choices by a pharmaceutical company directly leads to preventable harms to patients, and regulators should no longer tolerate this lack of scientific rigour. If we care about individual patient outcomes, we need **Scientific Regulators.**

Two regulatory pathways to approval were described, one based on **Population** outcomes (**Approval P**) and one based on **Individual** outcomes (**Approval I**). **Approval P** would be based on simple, fixed-dose, trial designs that use **Population** (average) patient outcomes for both benefits and harms. **Approval I** would be based on advanced trial designs where science-based dose titration algorithms are used to optimise the best **Individual** patient outcomes. **Approval I** will lead to **Personalised Dosing** and much better patient outcomes (our goal!), and the pharmaceutical industry can be encouraged to invest in the right trials to seek **Approval I** if, for example, it conferred a 5-year patent extension.

Two drug development strategies were introduced. **Strategy I** uses **Individual** outcomes to obtain **Approval I** for a smart science-based dose titration algorithm. It was shown how any dose titration algorithm to support **Approval I** would still produce tabular output for benefits and harms that is directly comparable to fixed-dose regimens. All stakeholders (regulators, patients, physicians, payers and the drug company) will help define what endpoints are important and work with analysts (like me) to design appropriate science-based dose titration algorithms. **Strategy P** uses **Population** outcomes to obtain **Approval P** for either **Fixed-Doses** (i.e. 1-2 doses) or a **Hybrid** dosing option, where a dose range in approved, with personalised dosing enabled via a simple dose titration algorithm. **Strategy P** is a simple drug development strategy, as it only requires a wide range of fixed-dose regimens to be studied throughout the whole drug development program.

The pharmaceutical industry also needs to carefully consider how **Strategy I** and **Strategy P** will deliver reimbursement and revenue. I discussed how modern value/outcome based pricing models are rightly aligning reimbursement with individual patient outcomes; individual patient outcomes are both the present and the future of reimbursement. The pharmaceutical industry must therefore embrace the challenge of employing **Personalised Dosing** to achieve the best outcome for each and every patient; it is what both patients and payers want, and it is therefore

in their commercial interests to produce a wide range of doses to enable each patient to find the dose that works best for them. Do this, and companies will maximise their chance of having a happy "customer" and a continued revenue source (Insulin glargine (brand name Lantus) yielded >$70 billion in lifetime sales because it did this very well). The benefits of a subscription based pricing model for drugs was also introduced, whereby reimbursement is fixed whilst the patient remains (happily) on the drug (i.e. pricing is independent of the dose). I also discussed how both the pharmaceutical industry and regulators should utilise modern technologies to deliver a much better product to the "user", the physician/patient. Current drug labels are verbose and not user friendly. In addition, post approval data that records patient engagement, adherence and PROs could, with permission, be collected across willing patients to enable multiple stakeholders (regulators, patient advocacy groups, payers) to further learn how different dosing regimens perform in a real world setting.

Finally, I introduced the value of using **adaptive randomisation** to most efficiently and ethically determining Population D-E-R relationships. This "learn as we go" approach starts with a randomisation schedule that supports a very wide dose range, and then "zeros in" on the optimal (most informative) doses based on updates to the randomisation allocations determined by a DMC. **Adaptive randomisation** is most valuable when our initial assumptions about the location and shape of the D-E-R relationships are imperfect. The ASTIN and AWARD-5 trials are two excellent examples that successfully employed **adaptive randomisation;** however the general scarcity of such trials for D-E-R trials would suggest the considerable advantages of these trials is currently not well understood. These adaptive trial designs can most quickly and cheaply quantify D-E-R relationships for both efficacy and safety, and hence should be a central component of any modern drug development program.

At this stage, I hope you are convinced of the "why" – the science tells us that we must seek to understand Population and Individual D-E-R relationships, and why we must always consider Personalised Dosing. Some of the "how" has also been addressed, including the general strategies to obtain Approval P based on Population (average) patient outcomes and Approval I based on individual patient outcomes.

What remains is the technical roadmap for the drug program and supporting trial designs; what does a company wishing to pursue **Approval P** and/or **Approval I** need to do, and how should the regulators work with them to ensure the evidence base for approval is as strong as it can be. These are the topics of the remainder of this book (in short, there are **no** technical barriers…we know how to do it!)

---

21 Technical Sections To Write

to do!

- Explain basic Clinical Trial Simulation (Simulation/Estimation * 1000)
- Explain key concepts around how we define/assess precision of DR
- Explain key concepts around how we define/assess bias of DR
- More advanced topics for optimal designs of DR, like D-optimality and V-optimality
- Compare optimal(adaptive) designs of D-E-R versus DR
- Introduce Planned Titration trials (a much better name for "forced" titration!)
- Explain the value of ER models for prediction (extrapolation to new regimens, pediatrics etc.)
- Explain the benefit of dose titration with respect to tolerance
- Have sections entitled:
    - Design and Analysis of Population D-E-R Trials
    - Design and Analysis of Individual D-E-R Trials
- Introduce technical code to fit D-E-R models? e.g. Bayesian analysis using HMC in STAN? Mention diffuse priors (better in Appendix?)
- Show case examples where drugs are individualised (to remove the mystery!)

22 Conclusions

I truly hope this text has provided you with further insights into how and why we need to better quantify D-E-R at both the **Population** and **Individual** level, and why we need to refocus drug development on individual patient outcomes.

If it has led to you using an improved design for your program and/or dose-ranging trials, please let me know. It would be nice to know it had a positive effect somewhere.

Equally, if you disagree with anything written herein, please also let me know. On many topics in drug development my views have evolved (changed!) over time through further learning and insights provided by friends/colleagues. For example, I initially believed that our goal was to find "the" optimal dose regimen; thus to determine the **Population** D-E-R relationships for both efficacy and safety/tolerability, and then select the dose which yielded the best "trade off" between these benefits and harms. It didn't occur to me that this is only the best strategy if we **must** only pick a single dose for approval. I am now much more aware that different patients need different doses, and simply forcing all patients to take the same dose will inevitably lead to unnecessary and **avoidable** under-dosing and over-dosing for many patients. I also now view safety data more through the lens of the patient, and see their suffering as a failure of the sponsor/team/myself to adequately tailor the dose for them.

If we are to make real progress in changing drug development for the better, we all need to be willing to "change our minds" as we transition from our old ways to a new, patient focused, drug development paradigm. I see this as a measure of the strength of character of an individual, and am very fond of this (mis) quote attributed to John Maynard Keynes:

> *"When the facts change, I change my mind. What do you do?"*

Although it is doubtful that Keynes actually said this, the quote is both excellent and clear. I would like to think my views on drug development have changed as I have acquired more "facts", and feel no shame in saying so.

I will leave you with two provocative options:

***Are you a part of the problem**, continuing to accept fixed-dose regimens because you are very familiar with "simple", even though this leads to worse patient outcomes?*

***Or are you part of the solution**, prepared to embrace **Personalised Dosing** because you truly care and want the best possible outcomes for our patients (even if the path looks a little "scary" and unfamiliar)?*

Our path to **Personalised Dosing** may not always be straightforward, and we will all need to better learn how to best implement it across all therapeutic areas. However fundamentally it is the right thing to do; be an advocate, speak up, and collectively our scientific arguments will prevail and convince the "simple" crowd to join us.

I am sure patients will appreciate our (imperfect) efforts to truly learn how best to deliver **Personalised Dosing** and achieve better patient outcomes; they are also entitled to be very angry with us when they suffer terribly because we persist with "simple" fixed-dose regimens. Remember, patients are not fields!

So let us move forward, and truly implement **Drug Development For Patients**.

23 Acknowledgements

I would like to thank the following people who have kindly offered valuable feedback and suggested improvements on early drafts of this work.

- Jill Feldman

- David Norris (Precision Methods)

- Ken Kowalski (Kowalski PMetrics)

(If you would like to be added above, just send me your comments/corrections/thoughts on how to improve this work!). I appreciate I am often too "blunt", so please tell me where I have gone too far!

I would also like to thank all my family, friends and colleagues who have encouraged me to write down my thoughts and/or improved my understanding of drug development and patient experiences.

There are many people who have inspired me. People like Lewis Sheiner, Radford Neil and Richard McElreath to name just a few. A common attribute these people share is their willingness to share their ideas and material to engage a wider audience and initiate debate. In a very small way, I am trying to follow in their footsteps.

Most importantly, I would like to thank my wife, Camilla, and kids Emma and Casper for politely listening to me as a talk about niche areas of drug development, and allowing me the time to spend on this project. Even though I know it is very far from perfect, I am happy that I am "having a go".

Best wishes,

Al

Glossary

ADME	Absorption, Distribution, Metabolism and Excretion
ANCOVA	Analysis of Covariance
ANOVA	Analysis of Variance
bid	twice daily (bis in die in Latin)
CPT	Clinical Pharmacology and Therapeutics (journal)
CTS	Clinical Trial Simulation
D-R	Dose-Response
D-E-R	Dose-Exposure-Response
FIM (with respect to trials)	First In Man
FIM (with respect to statistics)	Fisher Information Matrix
IIV	Inter-Individual Variability
INR	International Normalised Ratio
MAD	Multiple Ascending Dose (trial)
MBMA	Model-Based Meta-Analysis
MCP-MOD	Multiple Comparison Procedure - Modelling
MoA	Mechanism of Action
MTD	Maximum Tolerated Dose
N	Sample size (=number of trial participants)
NHST	Null Hypothesis Significance Testing
NSCLC	Non-Small Cell Lung Cancers
PD	Pharmacodynamic
PK	Pharmacokinetic
PPD	Price Per Dose
qd	once daily (quaque die in Latin)
RCT	Randomised Controlled Trial
RWE	Real World Evidence
SAD	Single Ascending Dose (trial)
SSE	Stochastic Simulation and Estimation
UX	User Experience

Contact

I hope you have found this book both enjoyable and thought-provoking, and that you look forward to seeing the remaining chapters.

I also hope it has inspired you to be an advocate for "Drug Development for Patients".

If you would like to discuss anything with me, I can be contacted at the email address below. It would be nice to hear your thoughts.

Best wishes, Al

1.   (1948) Streptomycin treatment of pulmonary tuberculosis. British Medical Journal 2:769–782

2.   Yoshioka A (1998) Use of randomisation in the medical research council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s. BMJ : British Medical Journal 317:1220–1223

3.   Maloney A, Rosenstock J, Fonseca V (2019) A Model-Based Meta-Analysis of 24 Antihyperglycemic Drugs for Type 2 Diabetes: Comparison of Treatment Effects at Therapeutic Doses. Clinical Pharmacology and Therapeutics 105:1213–1223. https://doi.org/10.1002/cpt.1307

4.   (1994) ICH E 4 Dose Response Information to Support Drug Registration

5.   Sheiner LB, Beal SL, Sambol NC (1989) Study designs for dose-ranging. Clinical Pharmacology & Therapeutics 46:63–77. https://doi.org/10.1038/clpt.1989.108

6.   Skrivanek Z, Gaydos BL, Chien JY, et al (2014) Dose-finding results in an adaptive, seamless, randomized trial of once-weekly dulaglutide combined with metformin in type 2 diabetes patients (AWARD-5). Diabetes, Obesity & Metabolism 16:748–756. https://doi.org/10.1111/dom.12305

7.   Woodcock J (2016) "Precision" drug development? Clinical pharmacology and therapeutics 99:152–154. https://doi.org/10.1002/cpt.255

8.   Collins FS, Varmus H (2015) A new initiative on precision medicine. The New England journal of medicine 372:793–795. https://doi.org/10.1056/NEJMp1500523

9.   Christakis P (2011) The birth of chemotherapy at yale. The Yale Journal of Biology and Medicine 84:169–172

10. Lesko LJ, Woodcock J (2004) Translation of pharmacogenomics and pharmacogenetics: a regulatory perspective. Nature Reviews Drug Discovery 3:763–769. https://doi.org/10.1038/nrd1499

11. Cook JC, Wu H, Aleo MD, Adkins K (2018) Principles of precision medicine and its application in toxicology. The Journal of Toxicological Sciences 43:565–577. https://doi.org/10.2131/jts.43.565

12. Powell JR, Cook J, Wang Y, et al (2021) Drug Dosing Recommendations for All Patients: A Roadmap for Change. Clinical Pharmacology and Therapeutics 109:65–72. https://doi.org/10.1002/cpt.1923

13. Maloney A (2021) Personalized Dosing = Approved Wide Dose Ranges + Dose Titration. Clinical Pharmacology and Therapeutics 109:566–567. https://doi.org/10.1002/cpt.1997

14. Hamberg A-K, Hellman J, Dahlberg J, et al (2015) A Bayesian decision support tool for efficient dose individualization of warfarin in adults and children. BMC medical informatics and decision making 15:7. https://doi.org/10.1186/s12911-014-0128-0

15. Kristensen NR, Agersø H (2022) Dose/exposure-response modeling in dose titration trials: Overcoming the titration paradox. CPT: pharmacometrics & systems pharmacology 11:1592–1603. https://doi.org/10.1002/psp4.12863

16. Ramanathan S, Jin F, Sharma S, Kearney BP (2016) Clinical Pharmacokinetic and Pharmacodynamic Profile of Idelalisib. Clinical Pharmacokinetics 55:33–45. https://doi.org/10.1007/s40262-015-0304-0

17. Li BT, Smit EF, Goto Y, et al (2022) Trastuzumab Deruxtecan in HER2-Mutant Non-Small-Cell Lung Cancer. The New England Journal of Medicine 386:241–251. https://doi.org/10.1056/NEJMoa2112431

18. Thomas N, Sweeney K, Somayaji V (2014) Meta-analysis of clinical dose–response in a large drug development portfolio. Statistics in Biopharmaceutical Research 6:302–317. https://doi.org/10.1080/19466315.2014.924876

19. Eichler H-G, Baird LG, Barker R, et al (2015) From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. Clinical pharmacology and therapeutics 97:234–246. https://doi.org/10.1002/cpt.59

20. Schnider TW, Minto CF, Filipovic M (2021) The Drug Titration Paradox: Correlation of More Drug With Less Effect in Clinical Data. Clinical Pharmacology & Therapeutics 110:401–408. https://doi.org/10.1002/cpt.2162

21. Maloney A (2017) A New Paradigm. "Learn - Learn More"; Dose-Exposure-Response at the Center of Drug Development and Regulatory Approval. Clinical pharmacology and therapeutics 102:942–950. https://doi.org/10.1002/cpt.710

22. Sheiner LB (1997) Learning versus confirming in clinical drug development*. Clinical Pharmacology & Therapeutics 61:275–291. https://doi.org/10.1016/S0009-9236(97)90160-0

23. Grieve AP, Krams M (2005) ASTIN: a Bayesian adaptive dose-response trial in acute stroke. Clinical Trials (London, England) 2: https://doi.org/10.1191/1740774505cn094oa

24. Maloney A, Karlsson MO, Simonsson USH (2007) Optimal adaptive design in clinical drug development: a simulation example. Journal of clinical pharmacology 47:1231–1243. https://doi.org/10.1177/0091270007308033