

A New Paradigm. “Learn – Learn More”; Dose-Exposure-Response at the Center of Drug Development and Regulatory Approval

Alan Maloney¹

In his seminal paper, Lewis Sheiner introduced the “Learning versus Confirming” paradigm. From that foundation, this work proposes why the precise estimation of the dose-exposure-response (D-E-R) for both efficacy and safety endpoints should be the ultimate goal for most drug development programs. The subsequent identification and approval of an optimal dose regimen range will provide a pragmatic framework for delivering personalized medicine based on dose titration for each and every patient.

Within this paper, I will use the term “utility” to reflect the composite of both benefit and risk, which themselves will be used interchangeably with efficacy and tolerability/safety respectively. The term “dose-exposure-response” (D-E-R) will be used to represent the sequence from dosing regimen (i.e., dose, frequency, route, etc) to drug exposure (i.e., typically actual or integrated drug concentrations in plasma (e.g., average concentration at steady state (C_{ss})) through to response (clinical measures of efficacy, tolerability, and safety).

The seminal paper by Sheiner¹ 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 20 years since he wrote this paper, two key components have become more central to modern drug development: (1) the importance of quantifying and predicting the safety and tolerability of different dosing regimens (in addition to efficacy); and (2) the goal of personalized medicine and its inter-relationship with wide interindividual variability (IIV) in response for efficacy, safety, and tolerability.

The former, the importance of being able to estimate how safety endpoints change as a function of drug exposure and patient covariates, was recognized by Sheiner.¹ Indeed 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”.

Today, “learning” is not simply about efficacy endpoints. 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 endpoints in 2017. We must design our studies to “learn” how safety endpoints change as a function of the drug regimen. That is, we must stop approaching safety analyses as a crude post-hoc exercise in integrating the (whatever is available) clinical trial data to one in which the design of the studies (i.e., the dose range studied + sample size) in the whole drug program are optimized to maximize the learning for safety endpoints. We must plan for these analyses and design our studies accordingly.

The second aspect is our goal of personalized medicine. In essence, personalized medicine is about getting the right drug and right dosing regimen (personalized dosing) for each patient.^{2,3} Woodcock² wrote: “The principal challenge in therapeutics is the variability of human responses to drugs, both for good and for ill. The ability to predict and consequently reduce this variation can significantly improve the benefit/risk balance of medicines.”

We must strive to better understand the shape of the D-E-R relationship at the population level (all patients), the subgroup level (all patients with covariates X and Y), and, most importantly, the patient level. As advocated by Woodcock,² using individual patient characteristics (e.g., sex, age, pharmacogenetics, etc.) to better tailor the drug regimen to the patient is commendable, strongly encouraged, and fully supported. This identification and use of patient-level characteristics to deliver better outcomes for patients is often termed “precision medicine,”⁴ and is likely to challenge current regulatory assessments based on population level inference.⁵ However, even within precision medicine, it is misguided to believe we will ever eliminate IIV, and the critical role played by dose; IIV in patient responses will remain ever present, and we must recognize this in both our approach to drug

¹Equation AB, Friggs Grand 4, 30275 Halmstad, Sweden. Correspondence: A Maloney (al_in_sweden@hotmail.com)

Received 10 February 2017; accepted 9 April 2017; advance online publication 17 April 2017. doi:10.1002/cpt.710

development and explicitly incorporate it as we look to best tailor medicines to patients. Two simple examples can be used to support this position.

First, there is warfarin. Warfarin is perhaps one of the most studied drugs, and Hamberg *et al.*⁶ reported a 10-fold range in the dose required for adequate anticoagulation in adults (assessed by the prothrombin time International Normalized Ratio). Although their tool can incorporate information on body weight, age, baseline International Normalized Ratio, target International Normalized Ratio, cytochrome P450 2C9, and VKORC1 genotypes, the authors advocate adaptive dosing to achieve the target International Normalized Ratio using Bayesian forecasting. That is, even after adjustments for multiple important covariates, there remains considerable IIV in the maintenance doses required, and individual titration is still important and necessary.

The second example is population pharmacokinetics (PKs). The PK/pharmacometric community has over 40 years of experience in estimating population PK models incorporating random effects that reflect IIV in key model parameters (e.g., clearance), and then identifying patient level covariates (such as weight, sex, age, renal function, etc) that influence these parameters. However, the reduction in the IIV after adjusting for covariates is often rather limited; considerable IIV still remains. This should not be surprising. Simply because two patients share some basic covariate data (e.g., two men who both weigh 80 kg, are 50 years old, and with normal renal function), there remains countless physiological, biological, pharmacological, and behavioral (e.g., compliance and diet) factors that could influence their own PK profiles (note: the ability of population PK models to better explain IIV in the future will depend on being able to better define and capture these more relevant covariates). Because pharmacodynamic (PD) variability to a given dose regimen is influenced by both PK variability (e.g., differences in concentration at steady state) and PD variability (i.e., PD IIV to the same exposure), the recognition and explicit incorporation of PD IIV into how we develop and approve all drugs must take center stage. IIV in response is our (ever-present) adversary.

It should be noted that the International Conference on Harmonization E4 guideline on Dose Response Information to Support Drug Registration,⁷ first published in 1994, highlighted a number of the points that will be covered herein. Some key passages are (the emphases are mine): “Knowledge of the relationships among dose, drug-concentration in blood, and clinical response (effectiveness and undesirable effects) is important for the safe and effective use of drugs in individual patients. This information can help identify an appropriate starting dose, the best way to adjust dosage to the needs of a particular patient, and a dose beyond which increases would be unlikely to provide added benefit or would produce unacceptable side effects.”

“Dose-response data for both beneficial and undesirable effects may provide information that allows **approval of a range of doses** that encompass an appropriate benefit-to-risk ratio.”

“In addition to carrying out formal dose-response studies, sponsors should examine the **entire database** for possible dose-response information... that can... identify a reasonable starting dose... and identify reasonable, response-guided titration steps,

and the interval at which they should be taken. . .with **appropriate adjustments for patient characteristics.**”

The highlighted text points to the centrality of the individual patient (not the whole patient population) and the use of dose titration via an approved dose range. Analysis of the “entire database” is also expected, although the document does not point out the information within this entire database is a function of the “input” study designs (quality of design => quality of inference).

In December 2014, the European Medicines Agency ran an excellent meeting entitled “Dose Finding Workshop.”⁸ Although there was universal agreement on the importance of “learning” via dose response studies, the reported stated: “The International Conference on Harmonization E4 acknowledges the importance of dose-response characterization and provides still valid recommendations. However, evidently it has not had the desired impact over the past 20 years probably due to insufficient specific guidance on dose-response requirements and methods.”

Thus, progress has been painfully slow (and indeed D-E-R continues to be seen as purely a phase II exercise that should be completed as hastily and cheaply as possible, so as to get one to two doses selected for phase III).

In addition to International Conference on Harmonization E4, the US Food and Drug Administration issued a guidance document on Exposure Response in 2003.⁹ It starts with an excellent statement: “Exposure-response information is at the heart of any determination of the safety and effectiveness of drugs.”

However, much of the discussion thereafter focuses on the use of exposure-response (E-R) as a mechanism for proposing dose adjustments in special populations (e.g., pediatric, geriatric, renal/hepatic impairment, etc), implicitly using “population average” changes in exposure to support dose regimen changes. For example, if the typical exposure in patients with moderate renal impairment was 100% higher compared with patients with normal renal function, a 50% reduction in dose may be proposed. However, to play devil’s advocate, perhaps the typical (efficacy/safety) half-maximal effective concentration (EC₅₀) for these patients is 100% higher, in which case the dose adjustment is wholly unnecessary (and actually detrimental). Thus, although PK arguments based on average exposure changes may seem reasonable, they both assume the underlying PD E-R relationships are identical in the two populations (i.e., “fixing” the PK is all that is necessary), and ignore the IIV in PK and PD. The key point here is that we need to adjust the dose to address IIV in PD response, and not simply/only for PK arguments based on average exposure.

The US Food and Drug Administration has recently published the Prescription Drug User Fee Act for fiscal years 2018–2022 (known as PDUFA VI),¹⁰ and it is very positive to note a number of important topics, consistent with the ideas herein, are proposed for further development and application. These include the design, analysis, and inference of D-E-R studies within a model-informed drug development setting, along with the use of complex, adaptive, Bayesian clinical trial designs, and the advancement of benefit-risk assessments in regulatory decision making.

This work posits that the use of one or two fixed dose regimens woefully fails to address PD IIV, leading to patients being

routinely underdosed and overdosed. This, in turn, leads to lower levels of efficacy and higher rates and severity of adverse events (that is, we are failing to deliver personalized medicine, as we are not getting the right dose to each and every patient). One consequence of overdosing is adverse drug reactions (ADRs). A European Commission report¹¹ concluded: (1) 0.12–0.22% of hospital admissions result in death due to an ADR corresponding to 100,800–197,000 deaths annually in the European Union; (2) 3–10% of hospital admissions are caused by ADRs corresponding to 2.5–8.4 million annually in the European Union; (3) 2.1–6.5% of hospitalized patients have an ADR, corresponding to 1.8–5.5 million annually in the European Union; €79 billion represents a reasonable estimate of the total societal cost of ADRs occurring in the European Union.

Dose titration will reduce both the rate and severity of any ADR. If we are capable of reducing the figures above, shouldn't we?

In addition, Radley *et al.*¹² reported 150 million off-label medications in the United States in 2001, and commented: "Most off-label drug mentions (73%; 95% confidence interval, 61–84%) had little or no scientific support."

The routine use of off-label drug use is the proverbial "elephant in the (regulators) room," and sits diametrically opposite to the fundamental principles of evidence-based medicine. For example, at the European Medicines Agency meeting in December 2014, one physician stated how he used most antipsychotics at doses substantially different from the approved doses. The report noted for this presentation: "This means that the B:R 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 2017? Has it ever been?

The paradigm outlined herein is relevant when the pharmacological goal is to reduce or manage current symptoms and/or to help prevent future complications, in contrast to life threatening situations in which the PD effect of drug treatment is time critical (e.g., heparin to treat a thromboembolism) and/or initiation of treatment with low doses may compromise outcomes (e.g., permit resistance development with antivirals and antibiotics). Although dose titration in life threatening situations may still be important and necessary, it is accepted that the approaches and ideas herein are specifically relevant to the more common situation of pharmacological interventions in which there is no imminent risk to the patient, and, hence, dose titration is reasonable.

This paper will first introduce individual and population D-E-Rs, and present why the search for an "optimal" dose is misguided. The significant limitations of studying a single dose in phase III are then covered. The "Learn – Learn More" paradigm is then introduced, from the philosophical foundations to the practical implementation. The critical role of the regulators is then discussed, and how the "Learn – Learn More" paradigm fits beautifully with the adaptive licensing/lifecycle benefit risk management approach to drug approval,^{13,14} and the regulators need to avoid excessive "risk aversion" in drug approval considerations.¹⁵ A number of issues relevant to the proposed paradigm

are highlighted in the "Additional Points to Consider" section, followed by the conclusion.

The goal of this work is to improve individual patient outcomes via a new paradigm for drug development and regulatory approval that permits a "start low, go slow" approach to dose titration over an approved "optimal dose regimen range." Science shows us the right path forward, but we will need determined pioneers to "grasp the nettle" and firmly put D-E-R at the center of drug development and approval, and deliver the regulatory changes needed to facilitate this new paradigm.

Individual and population dose-exposure-response

Central to this work is the distinction between a population D-E-R and the individual D-E-R for each patient. They are not the same.

Much of drug development involves studying cohorts of patients over a reasonably long period of time (e.g., 6 months), with each cohort receiving placebo or a different fixed dose regimen. The "average" or population D-E-R can then be estimated from combining all data using suitable D-E-R models, but importantly the results tell us nothing about individual D-E-R relationships, other than to confirm some individuals must have experienced some real change. Thus hypothetical, if dose regimen X yielded a population increase in heart rate of 2 bpm vs. placebo, then this may have been the result of all dose regimen X individuals experiencing a 2 bpm increase, or 10% of individuals experiencing a 20 bpm change, and 90% of individuals experiencing no change. Thus, the population D-E-R represents a simple "cross sectional" estimate of the average effect, but is incapable of telling us anything about the treatment effects in individual patients. **Figure 1** shows three different scenarios in which the population D-E-R is very similar, despite the underlying individual patient level D-E-R relationships being very different. Ideally, one would obtain individual patient level data over a wide range of doses to facilitate the estimation of individual D-E-R (in technical language, a D-E-R model with random effects that accurately capture IIV in the drug effect parameters; e.g., ED₅₀, Hill and maximum effect (E_{max})). Although study designs like this should be considered in some settings (e.g., like the heart rate example above), often the significant delay in observing the pharmacodynamic responses (e.g., HbA1c in diabetes) means that it may be difficult (or impossible) to consider such designs, and we are left with the basic population D-E-R models as our primary tool for dose selection. The main message from **Figure 1** is that knowing the population D-E-R does not provide us with any knowledge of how individual patients will respond to a given dose.

There is no one "Optimal" dose – patients are heterogeneous

With population D-E-R models available for multiple efficacy and safety parameters, there is often the discussion as the "optimal" dose regimen, with "optimal" being defined as the dose that would yield the highest utility (best benefit risk trade-off across all possible doses). Such approaches¹⁶ seem initially appealing, and would be perfect if we must pick a single dose. However, there is no such requirement.

Thus, the search for an "optimal" dose at the population level is a misguided goal, and indeed may not be optimal in any

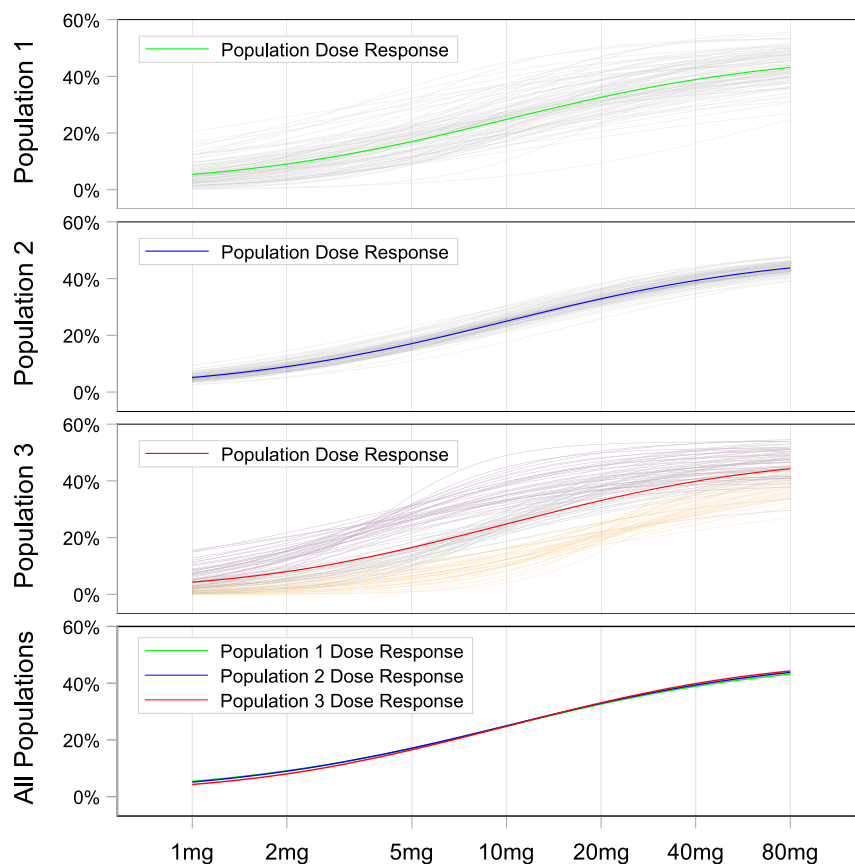


Figure 1 Population dose-response (D-R) curves (solid green/blue/red lines for populations 1, 2, and 3, respectively) shown alongside 100 random individual D-R curves. For population 3, there are 3 subpopulations (colored purple, gray, and orange). The three population D-R curves are shown superimposed in the last panel. [Color figure can be viewed at cpt-journal.com]

patient. This distinction is best illustrated using a therapeutic area much studied by Sheiner¹ – anesthesiology. Consider a new general anesthetic agent. If the dose is too low, the patient may be awake during surgery. If the dose is too high, the patient may experience respiratory depression or death. Thus, although we may indeed consider a single dose “optimal” if it minimizes (relative to all other doses) the fraction of patients who experience either of these negative effects, it is ridiculous to consider this a truly optimal use of the drug. Thankfully, anesthesiologists have utilized judicious dose titration strategies to tailor the dosing regimen to each patient (perhaps the perfect example of “personalized medicine”). Although the anesthesiologists may have the luxury of being able to rapidly determine the optimal dose for each patient based on the immediacy of the PD responses, the key point here is to recognize that each and every patient has their own D-E-R relationships for efficacy and safety, and, hence, utility function, in each and every therapeutic area. In addition, anesthesiology also tells us something further; despite the PK benefit of i.v. administration (e.g., no IIV in absorption as with oral agents), patients are highly heterogeneous with, for example, a very wide infusion dose range (>10-fold) recommended for maintenance of anesthesia with remifentanyl.¹⁷ We see similarly large IIVs for the mean daily doses of insulin in patients with type 2 diabetes.¹⁸ Thus, we must recognize that

patient heterogeneity is ever present, and wholly unaddressed with small two to three-fold “approved” dose ranges.

Thus, we should not ask what the optimal dose is, but rather, what is the distribution of individual optimal doses across the patient population. That is, whereas some patients may be optimally treated at 2 mg, other patients may be optimally treated at 20 mg. Thus, our goal is to allow the dose range of 2–20 mg to be available to the patients using dose titration. If (and when) we can better refine the dose regimens using patient covariates (e.g., sex, weight, renal function, comedications, pharmacogenetic information, etc), we clearly should.

Patients must be seen as a highly heterogeneous group, with any dosing regimen generating a wide range of systemic exposures [the dose-exposure (D-E) part] that are “married” to a wide range of individual responses to those exposures (the E-R part). Thus, we must see dose as just a very crude mechanism of delivering the drug to the site of action, and adjust it as needed to achieve the highest utility for each patient. No one dose is ever optimal.

What is wrong with looking at the benefit risk of a single phase III dose? Is this ethical drug development?

Without a well-defined D-E-R, we cannot say anything about whether a single dose represents the dose with the highest utility, even at the crude population level; perhaps lower or higher doses

would be significantly superior. In addition, we know nothing about the steepness of the D-E-R for the safety endpoints. Thus, for a drug with an IIV of 40% in clearance, there is about a 5-fold difference in steady state exposures across patients from the 2.5–97.5 percentiles. Therefore, while the “average” effect on a safety parameter may seem borderline acceptable for a given dose level, the patients with the higher exposures may be going into an area of the E-R with substantially higher risks than “average.” More importantly, this simplistic demonstration based on PK arguments ignores the even more important PD IIV, and, hence, patients who are particularly sensitive to the drug could avoid (or at least minimize the severity of) adverse events by starting at much lower doses and up-titrating.

It is worth commenting here on the (poor) practice of E-R modelling based on data from a single dose level (e.g., one exposure measure for one PD measure), in the hope the IIV in PK (like the 5-fold range discussed above) will reveal the true E-R. As an analyst, I know the partial confounding of exposure levels with covariate effects could be highly misleading [because we do not randomize patients to exposures, we lack balance (in a statistical sense) on the exposure axis]. For example, if reduced renal function is related to both higher exposures and an inferior efficacy outcome, the “overall” exposure response relationship may well be spurious (in addition to the general problem of being a mixture of the individual patient E-R relationships). Thus, the observed population E-R relationship may seem flatter or steeper than the true population E-R. At best, such an exercise is hypothesis generating. As a regulator looking at such an analysis for a safety endpoint, I would be forced to be highly skeptical. The scientific solution, at the design stage, is straightforward – study more dose levels and generate better input data for sound E-R analyses.

Currently, regulators are not tasked with assessing whether any dose regimen proposed by the sponsor is optimal in any sense, but rather to determine whether the benefit/risk (at the population level) is favorable. Thus, a single drug regimen that yields a positive individual utility (benefit/risk) in 55% of patients, but a negative individual utility in 45% of patients would be “approvable” based on the current remit, even though we may be doing net harm to 45% of patients. With only data from a single dose level, all safety analyses are intrinsically weak and limited, and, hence, we are also forcing our regulators to guess the actual risks of the regimen for our heterogeneous “real world” patient populations – this is not acceptable.

At this time, the reader may ask “But most drug programs study two dose levels in phase III, not one, so are the problems highlighted above still applicable?” In short, yes, especially when the higher dose is simply double the lower dose (a twofold dose range). Without using exposure data (i.e., just dose), the exercise is simply about fitting a linear D-R between the two treatment means, which is painfully crude and prone to error in both interpolation and extrapolation; such an analysis is quite useless. With exposure data, the limited dose range may yield only a small section of the E-R relationship, and any good analyst will desperately look to augment this weak analysis with phase II data considering a much wider dose/exposure range. Unfortunately, the small sample sizes in phase II will limit the actual information gained from such efforts,

especially for the analysis of binary (yes/no) safety endpoints. In summary, studying two dose levels in phase III is a very poor strategy if you consider the precise estimation of the D-E-R relationships for efficacy and safety as your primary goal (which I do).

As an aside, we should consider the ethics of phase III study design regarding dose levels and sample size. As argued herein, the whole drug development program, and, hence, the individual studies, should be (optimally) designed to best obtain sufficiently precise estimates of the D-E-R across multiple efficacy and safety endpoints to best facilitate an accurate benefit risk assessment, and not to demonstrate that the effect is not zero on a single primary efficacy endpoint in a single study [i.e., via the (mis-)use of power statements]. For example, for a diabetes drug entering phase III, a study may be “designed” to reject the hypothesis “Does the drug have no effect on HbA1c?” This is often akin to asking a physicist “Will the Sun rise tomorrow”? In phases I and II, the drug would have shown clear effects on fasting plasma glucose, postprandial glucose, and glycated hemoglobin (HbA1c). Thus, to determine the sample size to “disprove” a single, ridiculous null hypothesis is bordering on unethical. These patients, and this study, could be much better utilized by considering a very wide range of doses to either “rule out” or “rule in” different possible dose regimens. I would posit that it is unethical to recruit patients into studies and not maximize the information we learn from their participation (one of the many reasons why I advocate adaptive designs (see later for further discussion)).

The “Learn – Learn More” paradigm – from philosophy to implementation

The six principles that define the “Learn - Learn More” paradigms are:

1. IIV in efficacy, tolerability, and safety are always present.
2. Population D-E-R conducted across a wide dose range provides greater understanding of benefits/risks than studying any single dose.
3. Population D-E-R does not tell us how individuals will respond.
4. Individual titration across an approved dose range will improve outcomes for patients (higher benefits/lower risks).
5. Learning is about estimation, not significance testing.
6. Adaptive designs (e.g., changing the randomization allocations based on the accruing data) are more informative, and, hence, more ethical than fixed designs.

Points 1–4 have been discussed previously, but perhaps it is useful to provide an illustration of what this would look like in practice. In **Figure 2**, the D-R for three endpoints (one efficacy, one tolerability, and one safety) are shown for a hypothetical drug program that studied 2,500 patients using a standard parallel group study (500 patients at each of placebo, 40 mg, and 80 mg, and 200 patients at each of 1 mg, 2 mg, 5 mg, 10 mg, and 20 mg). The endpoints are all binary, and the population D-R and 90% prediction interval for the change vs. placebo is shown (full details of the simulation and estimation methods are provided in **Supplementary Material S1**). The figure also illustrates a

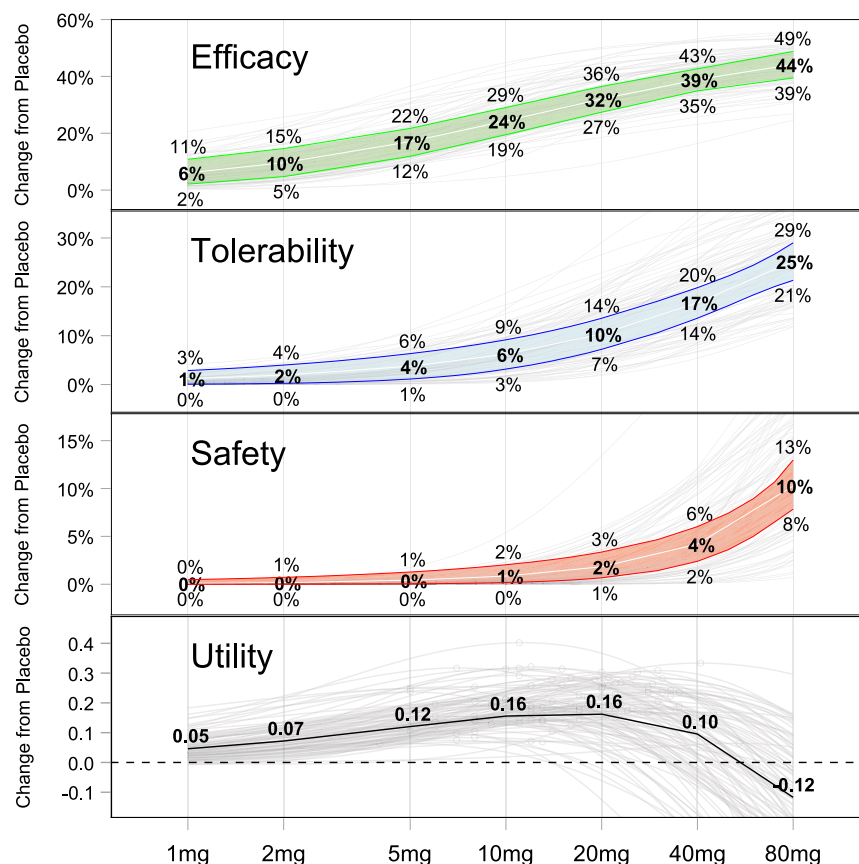


Figure 2 The estimated population dose-response (D-R) and 90% prediction interval (white lines and shaded regions) for an efficacy, tolerability, and safety endpoint are shown in the top three panels. The light gray lines are 100 random individual D-R curves; these are typically unobserved but which, 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. [Color figure can be viewed at cpt-journal.com]

random sample of 100 individual D-R relationships (gray lines), to highlight the differences between individuals and the population average response. Finally, the last panel shows a utility (benefit/risk) assessment using a simple utility function, in which the weighting is 1:1:3 for efficacy/tolerability/safety, respectively. Thus, a 1% increase in efficacy is considered “equivalent” to a 1% increase in tolerability, but a 1% increase in safety would need a 3% increase in efficacy to be considered “equivalent.” This type of clinical utility index (also called multicriteria decision analysis) is one method of quantitatively balancing benefits vs. risks, as well as each reviewer simply moving their eyes across the dose axis to determine their own “preferred” drug profile. Importantly, the final figure shows the utility curves for our 100 random individuals, along with their maxima (circles). Of note, these individual utility curves have maxima from below 5 mg to above 40 mg. Thus, whereas the population utility curve might suggest 10 mg and 20 mg as the “optimal” doses, the individual utility curves suggest we can do much better. Indeed, this simple example crudely assumes all patients have the same utility function (i.e., they would all use identical weighting function across the three endpoints, which is unrealistic.¹⁹ Based on these population results, individual titration from 2 mg up to 20 mg might be reasonable for approval. A future study could consider “difficult to treat”

patients, who tolerate 20 mg, but do not reach a sufficient efficacy response at that dose. These patients could be randomized to either dose escalation (e.g., to 40 mg or higher) or placebo (i.e., stay on 20 mg). These data would then support or refute the value of additional titration beyond 20 mg in such patients.

A number of points are worthy of discussion:

1. All endpoints were binary, and, thus, arguably we are using the least informative type of data in this example (cf. normal type data).
2. The precision of the D-R relationships is high, and the use of a wide dose range permits a much clearer understanding of the risks at, for example, 10 mg and 20 mg, than if only these doses had been studied.
3. It took the computer <20 s to fit each nonlinear D-R model using advance methods of estimation that accurately capture the uncertainty in the model predictions (see **Supplementary Material S1**). Thus, we can routinely apply D-R analysis methods as our primary analyses instead of simplistic and naïve “by study/by dose” methods (e.g., analysis of covariance for each study).
4. The extension to multiple efficacy and safety endpoints is straightforward, and permits a clear overview of all population dose responses.

5. The fixed design/randomization allocation used was not optimized in any way. Thus, better designs (yielding lower uncertainty/higher precision across the three endpoints) could have been found, thereby learning more than we actually did (higher precision = lower uncertainty), or, equivalently, reducing the total sample size to achieve the same overall precision (see refs. 20, 21 for examples of optimal design for D-E-R modeling).

In all of the above, it is worth noting that the paradigm is one of estimation, not of significance testing. The “significance testing” approach to design is concerned with being able to reject statements such as “the drug effect is not zero,” whereas the “estimation” approach is concerned with quantifying the magnitude of the positive and negative effects of a drug (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, is unhelpful. As shown in **Figure 2**, an estimation framework can provide everything needed for a rational approach to dose range selection (no *P* values required).

Finally, although the main thrust of this paper is not to explain the value of adaptive designs, it would be remiss to not mention why they fit so perfectly from a scientific, ethical, and cost/time perspective, and, hence, why they should be seen as an integral part of the “Learn – Learn More” paradigm. To be precise, the adaptive element is the use of adaptive randomization, whereby the randomization allocations (percentage of patients randomized to each dose level) are continually optimized as the location and shape of the D-E-R relationships (for efficacy and safety) are “learned” from the accruing data (note: the study can remain double/triple blinded). Thus, while the starting design (dose range and initial randomization allocation) should be optimized prior to the study start (using all available knowledge at that time), it is judicious to use the accruing data to update the randomization allocations to maximize the information gained from each patient (and wholly avoid, for example, the situation in which most or all doses are at the bottom or top of the D-E-R curve). This will always be more efficient than a fixed design, and most valuable when the new data is least consistent with the prior expectations.²² To be direct, running D-E-R studies as fixed designs is to be deliberately inefficient. By ensuring each and every patient contributes most to our understanding of the D-E-R using adaptive randomization, it is not only more efficient (in cost/time/patients), it is “more” ethical, because we are not exposing patients to dose levels that contribute minimally to our understanding (note: the study can also have built-in stopping rules for futility to ensure recruitment is stopped if the data suggest a positive outcome is very unlikely).

To conclude, if the six principles are agreed and supported, we have the methods and tools to implement them in practice.

The three components that need to change to facilitate this new paradigm

Figure 3 gives an overview of the “Learn – Learn More” paradigm embedded into the drug development and regulatory review/approval framework. The continued learning and

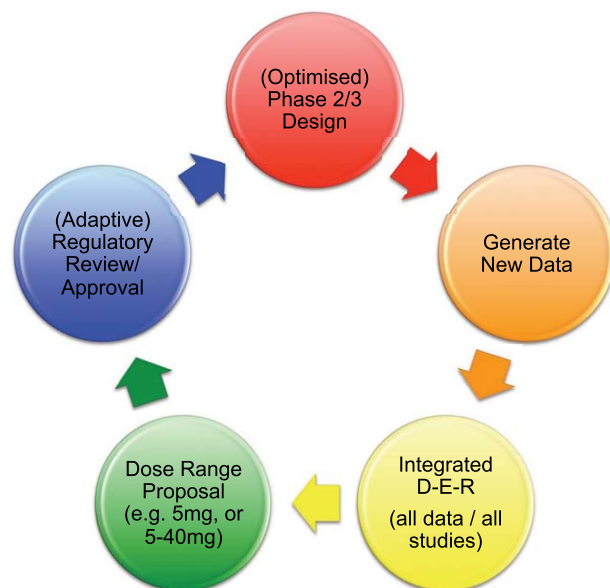


Figure 3 The “Learn – Learn More” paradigm for drug development. The learning cycle consists of design clinical studies for the planned dose-exposure-response (D-E-R) analysis, then collecting the new data, and then performing the analyses. When the evidence base is sufficiently strong, one or more dose levels may be proposed for approval. An adaptive regulatory approach naturally facilitates extensions or amendments to the approved dosing regimens based on the accruing evidence base, and “feeds back” into the ongoing clinical development program (that is, the whole process is cyclic). [Color figure can be viewed at cpt-journal.com]

refinement of the D-E-R relationships for efficacy and safety generate an ever stronger evidence base, thus permitting an adaptive regulatory licensing strategy, whereby lower doses may be initially approved, followed by higher doses as the evidence base becomes stronger. To facilitate this paradigm, at least three factors need to be fundamentally changed.

First, the integrated D-E-R analyses needs to be considered the primary analysis. By integrated, this means “across all late phase studies” and across all doses tested. Although it will always be wise to review each study separately, it is essential to combine data across all studies to determine the best and most precise overall estimates (and uncertainty) of the D-E-R for each efficacy, tolerability, and safety endpoint. At the most basic level, this could be a simple D-R analysis of each endpoint independently, through to more complex D-E-R longitudinal models (potentially across multiple correlated endpoints) incorporating, for example, disease progression, covariate effects, and alternative methods for missing data imputation.

Second, when the integrated D-E-R is considered the definitive key analysis, the studies should be designed accordingly, with wide dose ranges studied in all studies. Phase III studies should, in general, be essentially large phase II studies, and designed to obtain sufficient precision on the treatment effects for key efficacy and safety endpoints, rather than designed to “reject” a hypothesis that the treatment effect for a single efficacy endpoint is not zero (by focusing on the power of a single efficacy endpoint, current phase III designs woefully fail to “confirm” benefit

risk in any way). Importantly, the separation between phase II and phase III need no longer exist, because the studies could be essentially combined, with the remit simply to acquire sufficient data across a broad range of doses to facilitate an optimal dose regimen range for approval. To understand IIV in response, randomized crossover studies (such as balanced incomplete block designs), forced titration, and flexible titration studies can augment fixed, parallel group designs.

Third, the regulators must drive this change. If the regulatory framework is changed such that the two steps above are placed central to the drug approval process, the drug industry will change. By itself, the drug industry is very unlikely to change, because the upside for the industry (more patients staying on treatment for longer, better overall outcome for patients, potentially shorter drug development programs and earlier to market, lower adverse event rates, etc.) will be lost to the commercial demands and operational (“same as last time”) simplicity of the current (flawed) approach.

The last point is, by far, the most difficult to achieve. Although I am wholly optimistic good scientists will embrace the scientific foundations of the “Learn – Learn More” paradigm at a personal level, it will require significant commitment from senior regulatory stakeholders to deliver this vision. As science moves forward, so must our regulatory frameworks.

Adaptive licensing / lifecycle benefit risk management

The proposed paradigm works beautifully with the concepts of adaptive licensing/lifecycle benefit risk management. That is, rather than see the regulatory process as a single approve/reject exercise, to see it as one that is continually refined in light of an ever-stronger evidence base (data/results). Technically, every new patient helps improve the precision of the D-E-R models, to the point where a given dose or dose range may be considered suitable (or not) for approval. Depending on the indication and/or unmet need (e.g., orphan indications), regulators may be prepared to initially grant early approval even when the D-E-R predictions are associated with high uncertainty (that is, wide prediction intervals), because the early results may be highly compelling and/or the need pressing. For example, the initial European Union approval of Glybera (a gene therapy treatment that compensates for lipoprotein lipase deficiency) was based on 27 patients.²³ Clearly, with 27 patients the uncertainty across all efficacy and safety endpoints would be extremely high, however, initial approval may be reasonable, with the caveat that more data must be generated. There will always be a balance between higher sample size (to reduce bias and uncertainty) and regulator risk,²⁴ and it is for all parties (regulators, patient groups, therapeutic experts, industry, etc) to define where the thresholds are for each indication.

In addition, it seems natural to approve lower dose levels initially, because these will be least likely to cause harm. At this stage, the target population may also be restricted. As the evidence base strengthens, the greater precision for the D-E-R predictions could support the approval of higher doses (and/or the broadening of the target population). The product label could then be updated to reflect the dose range available, along with a recommended titration strategy for physicians to use (e.g., the

starting dose, and the minimum duration of exposure; e.g., 1 week, 1 month, etc.) before any subsequent up-titration, based on both PK and PD knowledge of temporal delays in observing full treatment effects. Knowledge of interoccasion variability could also potentially augment the titration strategy. For example, dose titration could be proposed if a patient with diabetes had (after treatment) a fasting plasma glucose above a given threshold, however, because interoccasion variability may be expected for fasting plasma glucose, it may be prudent to require multiple fasting plasma glucose samples (e.g., taken more than 1 week apart) to be consistently above the given threshold before up-titration, to ensure patients are not inappropriately up-titrated [based on a single (randomly high) sample].

Ideally, regulators and sponsors would communicate early on the relative importance of endpoints for the D-E-R analyses, such that the process is transparent for all. The role of clinical utility index/multicriteria decision analysis offers a quantitative approach for aggregating the D-E-R results across endpoints to assess benefit-risk. Indeed, one could imagine that every 3–6 months the accrued evidence base is aggregated to (re-)evaluate the D-E-R results, to determine if the approval status for any/all doses should be revised. For sponsors, a candidate drug failing to demonstrate sufficiently positive benefit risk could be swiftly terminated, whereas a highly positive benefit-risk profile should be given greater resources (e.g., faster recruitment) and considered for early approval. This achieves the twin goals of getting patient access to innovative medicines quicker, while helping the sponsor best utilize their finite resources across their portfolio (e.g., terminating drugs early when the probability of reaching key goals is low, while fast tracking their most promising drug candidates).

Additional points to consider

Supplementary Material S2^{25–33} briefly discusses a number of topics not mentioned elsewhere, but relevant to the application of “Learn – Learn More” in practice (e.g., discussing the “where and how” rather than the “why” of the main paper). This additional material is organized into the following sections:

- Applicable therapeutic areas
- Analysis methods: What D-E-R model?
- Covariate analysis, not subgroup analysis
- Use of external data and active controls
- Pediatric drug development and off-label drug use – moving from a bad system to a good one
- Dose individualization – progress is being made.

CONCLUSION

This paper has briefly outlined why the precise estimation of the D-E-R for both efficacy and safety endpoints should be the ultimate goal for most drug development programs. By identify an optimal dose regimen range, patients can use dose titration to achieve their own maximum utility (benefit vs. risk). This will lead to improved individual patient outcomes. Our extensive knowledge of IIV in both PK and PD (i.e., science) tells us that for drug dosing regimens, one size does not fit all. We must stop pretending that is does, and recognize the need for change.

Drug development and approval should be a scientific exercise in “Learn – Learn-More.” The methods to design and analysis D-E-R studies for both efficacy and safety are well understood; no technical barriers exist.

In 1997, Sheiner¹ wrote: “I and others have recently advocated increased attention to science (learning) in clinical drug development. Some of us have lamented the lack of scientific orientation of clinical drug development.”

Because most drug development programs I see today consist of one or two fixed dose regimens in phase III studies designed around significance testing for a single primary efficacy endpoint, I am pained to agree with this statement today. Equally, I remain convinced that a more scientific approach to drug development and approval (like the one outlined herein) will ultimately prevail. It is hoped this paper will stimulate the discussion, refinement, and, most importantly, the implementation of “Dose-Exposure-Response at the center of drug development and regulatory approval.”

To expedite the transition to this new paradigm, it is highly likely that regulatory agencies will play a critical role, but will need to update their legislative remit to do so.

Additional Supporting Information may be found in the online version of this article.

ACKNOWLEDGMENTS

The author kindly thanks Janet Wade and Kajsa Hamberg for their valuable comments and advice on an early draft of this work, and the reviewers who suggested a number of very useful additions and clarifications that improved this paper.

CONFLICT OF INTEREST

The author declared no conflict of interest.

© 2017 American Society for Clinical Pharmacology and Therapeutics

1. Sheiner, L.B. Learning versus confirming in clinical drug development. *Clin. Pharmacol. Ther.* **61**, 275–291 (1997).
2. Woodcock, J. “Precision” drug development? *Clin. Pharmacol. Ther.* **99**, 152–154 (2016).
3. Morse, B.L. & Kim, R.B. Is personalized medicine a dream or a reality? *Crit. Rev. Clin. Lab. Sci.* **52**, 1–11 (2015).
4. Collins, F.S. & Varmus, H. A new initiative on precision medicine. *N. Engl. J. Med.* **372**, 793–795 (2015).
5. Breckenridge, A., Eichler, H.G. & Jarow, J.P. Precision medicine and the changing role of regulatory agencies. *Nat. Rev. Drug Discov.* **15**, 805–806 (2016).
6. Hamberg, A.K., Hellman, J., Dahlberg, J., Jonsson, E.N. & Wadelius, M. A Bayesian decision support tool for efficient dose individualization of warfarin in adults and children. *BMC Med. Inform. Decis. Mak.* **15**, 7 (2015).
7. ICH Topic E4: dose response information to support drug registration. <http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E4/Step4/E4_Guideline.pdf> (1994).
8. European Medicines Agency - Report from Dose Finding Workshop, London, 4–5 December 2014. <http://www.ema.europa.eu/docs/en_GB/document_library/Report/2015/04/WC500185864.pdf> (2014).
9. US Food and Drug Administration (FDA) guidance for industry: exposure-response relationships — study design, data analysis, and regulatory applications. <<http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm072109.pdf>> (2003).
10. U.S. Food and Drug Administration. PDUFA reauthorization performance goals and procedures fiscal years 2018 through 2022

(known as PDUFA VI). <<https://www.fda.gov/downloads/ForIndustry/UserFees/PrescriptionDrugUserFee/UCM511438.pdf>> (2016).

11. European Commission. Proposal for a regulation amending, as regards pharmacovigilance of medicinal products for human use. Regulation (EC) No 726/2004. Impact assessment. <http://ec.europa.eu/health/files/pharmacos/pharmpack_12_2008/pharmacovigilance-ia-vol1_en.pdf> (2008).
12. Radley, D.C., Finkelstein, S.N. & Stafford, R.S. Off-label prescribing among office-based physicians. *Arch. Intern. Med.* **166**, 1021–1026 (2006).
13. Eichler, H.G. et al. From adaptive licensing to adaptive pathways: delivering a flexible life-span approach to bring new drugs to patients. *Clin. Pharmacol. Ther.* **97**, 234–246 (2015).
14. Baird, L.G. et al. Accelerated access to innovative medicines for patients in need. *Clin. Pharmacol. Ther.* **96**, 559–571 (2014).
15. Eichler, H.G. et al. The risks of risk aversion in drug regulation. *Nat. Rev. Drug Discov.* **12**, 907–916 (2013).
16. Kodell, R.L. & Chen, J.J. Quantitative benefit-risk analysis for evaluating drug therapies. *J. Biopharm. Stat.* **23**, 231–238 (2013).
17. Ultiva. Abbott Laboratories Abbott Laboratories, North Chicago, IL, USA. <http://www.accessdata.fda.gov/drugsatfda_docs/label/2004/20630se5-005_ultiva_lbl.pdf> (2004).
18. Eby, E.L., Van Brunt, K., Brusko, C., Curtis, B. & Lage, M.J. Insulin dosing and outcomes among commercially insured patients with type 2 diabetes in the United States. *Clin. Ther.* **37**, 2297–2308 (2015).
19. Postmus, D. et al. Incorporating patient preferences into drug development and regulatory decision making: results from a quantitative pilot study with cancer patients, carers, and regulators. *Clin. Pharmacol. Ther.* **99**, 548–554 (2016).
20. Maloney, A. et al. An example of optimal phase II design for exposure response modelling. *J. Pharmacokinet. Pharmacodyn.* **37**, 475–491 (2010).
21. Maloney, A., Simonsson, U.S. & Schaddelee, M. D optimal designs for three Poisson dose-response models. *J. Pharmacokinet. Pharmacodyn.* **40**, 201–211 (2013).
22. Maloney, A., Karlsson, M.O. & Simonsson, U.S. Optimal adaptive design in clinical drug development: a simulation example. *J. Clin. Pharmacol.* **47**, 1231–1243 (2007).
23. Melchiorri, D. et al. Regulatory evaluation of Glybera in Europe - two committees, one mission. *Nat. Rev. Drug Discov.* **12**, 719 (2013).
24. Zineh, I. & Woodcock, J. Clinical pharmacology and the catalysis of regulatory science: opportunities for the advancement of drug development and evaluation. *Clin. Pharmacol. Ther.* **93**, 515–525 (2013).
25. Manolis, E., Rohou, S., Hemmings, R., Salmonson, T., Karlsson, M. & Milligan, P.A. The role of modeling and simulation in development and registration of medicinal products: output from the EFPIA/EMA modeling and simulation workshop. *CPT Pharmacometrics Syst. Pharmacol.* **2**, e31 (2013).
26. Hemmings, R. An overview of statistical and regulatory issues in the planning, analysis, and interpretation of subgroup analyses in confirmatory clinical trials. *J. Biopharm. Stat.* **24**, 4–18 (2014).
27. Kennedy, M.J. Characterizing the impact of development on the dose-exposure-response relationship: challenge or opportunity? *J. Pediatr. Pharmacol. Ther.* **19**, 260–261 (2014).
28. Mehrotra, N. et al. Role of quantitative clinical pharmacology in pediatric approval and labeling. *Drug Metab. Dispos.* **44**, 924–933 (2016).
29. Mould, D.R., D’Haens, G. & Upton, R.N. Clinical decision support tools: the evolution of a revolution. *Clin. Pharmacol. Ther.* **99**, 405–418 (2016).
30. Bloemen, S., Hemker, H.C. & Al Dieri, R. Large inter-individual variation of the pharmacodynamic effect of anticoagulant drugs on thrombin generation. *Haematologica* **98**, 549–554 (2013).
31. van Dijkman, S.C., Alvarez-Jimenez, R., Danhof, M. & Della Pasqua, O. Pharmacotherapy in pediatric epilepsies: from trial and error to rational drug and dose selection - a long way to go. *Expert Opin. Drug Metab. Toxicol.* **12**, 1143–1156 (2016).
32. Calvo, E., Walko, C., Dees, E.C. & Valenzuela, B. Pharmacogenomics, pharmacokinetics, and pharmacodynamics in the era of targeted therapies. *Am. Soc. Clin. Oncol. Educ. Book* **35**, e175–e184 (2016).
33. Staiger, H., Schaeffeler, E., Schwab, M. & Haring, H.U. Pharmacogenetics: implications for modern type 2 diabetes therapy. *Rev. Diabet. Stud.* **12**, 363–376 (2015).