 doi:10.20944/preprints202009.0481.v1

Essay

Mechanistic Research for the Student or Educator †

Rehana Leak

Duquesne University; leakr@duq.edu

† Part I of II

Abstract: Many discoveries in the biological sciences have emerged from observational studies, but student researchers also need to learn how to design experiments that distinguish correlation from causation. For example, identifying the physiological mechanism of action of drugs with therapeutic potential requires the establishment of causal links. Only by specifically interfering with the purported mechanisms of action of a drug can the researcher determine how the drug "causes" its physiological effects. Typically, pharmacological or genetic approaches are employed to modify the expression and/or activity of the biological drug target or downstream pathways, to test if the salutary properties of the drug are thereby abolished. However, experimental techniques have caveats that tend to be underappreciated, particularly for the newer methods. In this two-part series, the caveats and strengths of mechanistic preclinical research are described, using the intuitive example of pharmaceutical drug testing in experimental models of human diseases. This series is not intended to tackle the perpetual clash between the frequentist approach to statistics and other schools of thought. Rather, Part I focuses on technical practicalities and common pitfalls of cellular and animal models designed for drug testing, and Part II describes in simple terms how to leverage a full-factorial three-way ANOVA, to test for causality in the link between drug-induced activation (or inhibition) of a biological target and therapeutic outcomes. Upon completion of this series, the student is expected to appreciate the strengths as well as limitations of mechanistic research and to avoid some of its pitfalls.

Keywords: mechanistic; hypothesis; physiology; biology; pharmaceutical; biomedicine; preclinical

Introduction

Physiology and pathology are both systemic and reductionistic disciplines, often operating at the level of proteins or their substructures, but also at the higher levels of cell death, organ toxicity, and organism mortality. Using the analogy of a child taking apart a watch and studying the individual tiny pieces, but failing in how to tell time, Bechtel argued, "the full causal account of a mechanism must include the causal processes at all levels or it will not be able to account fully for the phenomena" (1). He further commented, "the mechanistic framework accommodates this complementary relationship between reductionistic and systems approaches" (1). Mechanistic reasoning in physiology and pathology has gained considerable traction with the emergence of "hypothesis testing", a relatively intuitive process for beginner students (2). However, hypothesis testing is an exercise fraught with caveats, which students may not appreciate until time, effort, and monies are squandered. The goal of Part I in this series is to tackle this gap in graduate education.

STEM educators are responsible for training students in the rigorous conduct, evaluation, and analyses of hypothesis testing in mechanistic research. However, the PubMed search terms "(education) AND (mechanistic) AND (hypothesis)" or the alternate search terms "(teaching) AND (mechanistic) AND (hypothesis)" do not capture publications on teaching students how to identify biological mechanisms underlying natural phenomena or pharmaceutical drug action. Thus, this two-part series describes how to test two commonly-encountered hypotheses, using preclinical drug testing as a straightforward example, and the common pitfalls associated with this scholarly pursuit:

Hypothesis 1: Drug D (or any kind of intervention) prevents the effects of model disease M by upregulating the function of molecule P

Hypothesis 2: Drug D (or any kind of intervention) prevents the effects of model disease M by downregulating the function of molecule P

To test these hypotheses, it is necessary to determine if the therapeutic effects of the drug can be abolished (or at least attenuated) when the proposed mechanism of action has been interfered with. In this example, the null hypothesis states that drug D continues to mitigate the toxic effects of model disease M, even when the function of protein P has been interfered with. However, rejection or retention of the null hypothesis will not be argued further in this series (3). Rather, the focus is on the test hypothesis, as in a modern grant proposal.

The reader may wonder if they really need a mechanistic hypothesis to commence their laboratory work. The short answer is, "Of course not". The beginner does not even need a non-mechanistic hypothesis—they can perform a study to simply observe and describe a biological phenomenon. Unlike mechanistic work, descriptive and observational studies are not dependent upon crafting any *a priori* hypothesis. Although this type of descriptive research may be viewed less favorably by reviewers of papers and grant applications, it can be less biased than hypothetico-deductive research, by virtue of its exploratory and open-ended nature (4, 5). The beginner may find that descriptive research is essential in the initial phases of the project, and without it, they might not be able to formulate a rational mechanistic hypothesis for follow-up work. Descriptive and hypothesis-driven research are complementary, rather than mutually exclusive, as they serve to propel each other forward (6).

This Educational Series Does Not Replace Textbooks on Experimental Design or Statistics

Budding researchers will need to learn why results are not necessarily biologically meaningful because there is a statistically significant effect. For eye-opening discussions of the history behind hypothesis testing and the challenges in determining "statistical significance" vis-à-vis "practical importance", the reader is guided to the writings of statisticians (3, 7-11). This series assumes forehand knowledge of fundamental statistical tenets, such as randomization of sampling, *a priori* specification of exclusion criteria, power analyses, assumptions of normality and homoscedasticity by ANOVAs, *etc*. Also beyond the scope of this series are the historical origins of mechanistic thought in physiology, originating partly in Claude Bernard's work, the distinction between mechanistic, mathematical, or narrative thinking, and some of the roots of the scientific process in Hellenic and Hebraic traditions (12).

Flaws (and Strengths) of Disease Modeling

Humans and wild animals are difficult to model in the lab because they display greater genetic diversity than animal colonies at research institutions and do not live in tightly controlled temperature conditions or with clean, *ad libitum* food and water. Inbreeding and substrain effects in animal colonies have long plagued mechanistic research (13-15). The fallacy of "standardization" of research models (16, 17) is not widely discussed in the biological, pharmaceutical, or biomedical sciences. Simply put, the more often a physiological effect is observed across substrains, species, and technically heterogeneous models, the more generalizable the effect is, and the more likely to translate to humans.

The evolutionary divergence of rodent and hominid ancestors does not necessarily impair the study of physiological processes in the laboratory. It depends on the specific process under examination. Although the cerebral cortices of these two groups diverged in critical ways, the common refrain that cultured cells and rodent animals are too different from human bodies to be useful is too simpleminded. Rather than blaming evolutionary divergence, students should be aware that we not understand the etiology of many human diseases, and it is therefore difficult to know what to "induce" in the rodent (or even the nonhuman primate) to feasibly model the human

 doi:10.20944/preprints202009.0481.v1

condition. In simple words, the fault may lie in what we fail to do to the animals to model the disease, not in animal physiology *per se*. Some exceptions to this might involve inherited or genetic conditions, as DNA sequences are easy to technically assess and manipulate. For example, investigators have leveraged disease-causing mutations in human induced pluripotent stem cells in the hope of avoiding the caveats of working on cells from other species (18, 19). Although these studies have the limitation of being conducted *in vitro*, cell culture also has the advantage of affording exquisite control over all independent variables. Preclinical cellular and animal models occupy an essential place in mechanistic research, as they allow for the appropriate randomization and control over key environmental and genetic variables. The latter requirements are difficult to attain or unethical in human subject research. However, it is important for the student to understand that "a model is a model". By very definition, it can never be the human disease itself, even if enthusiasm runs high when it is introduced. In sum, fully generalizable or clinically translatable observations are difficult to attain, especially when financial, technical, or human resources are limited in a lab. It is therefore wise to adopt a skeptical but also forgiving attitude.

Misapplying the Dogma of 'Form Follows Function'

In both descriptive and mechanistic studies, multiple technical approaches should be used confirm the data from different angles to ensure generalizability (see above), including independent measurements of anatomical structure and physiological function. Although structure clearly reflects function, this doctrine can be misapplied, because technical assays are not all-encompassing and only measure *specific aspects* of structure or function. For "structural" measurements of cellular viability, cell numbers are typically counted by microscopy. For functional measurements of cellular viability, it is common to measure the energy capacity of the cell (20). However, the therapeutic intervention might stimulate energy production per cell; the conventional assumption that bioenergetics are in proportion to cell counts is then invalid. Conversely, cell counts might show that the drug therapy prevents cell loss, but none of the preserved cells might be functioning properly—perhaps the cells are structurally present, but all might be incapacitated by low energy levels.

For *in vivo* studies involving drug-testing, functional assessments of measurement outcomes are often used to complement the structural data. For example, a battery of behavioral and electrophysiological tests can rule out the possibility that brain cells are structurally protected by the intervention at the histological level, but unable to fulfill their normal physiological roles. Although many believe that structural changes need to be profound to be manifested as observable changes in organismal behavior, cell loss probably does not need to be dramatic to generate a clinical syndrome, as was once held (21). In addition, control over behavior is not mediated by unitary brain regions as is sometimes presented, but, as argued by electrophysiologists back in 1970: "the available evidence indicates that the substrate for control of behavior must be located within a variety of sites in the central nervous system" (22).

Finally, it is not widely appreciated that biochemical, behavioral, and histological assays can be confounded by the stress and anxiety that the presence of the investigator provokes in the animal, an effect that may be greater for male investigators (23). To combat this, frequent animal handling is recommended to temper activation of the hypothalamic-pituitary-adrenal axis (24).

Insensitive Measurement Tools Readily Beget False Outcomes

Assay validation is an important first step in any preclinical research, as insensitivity of measurement tools can result in Type II errors (*i.e.*, false negatives). When relying on viability assays on cells grown in a dish, seeding cells in the wells of the dishes at a wide range of increasing densities can help ensure that the viability assay resolves changes in cell numbers accurately (**Figure 1**). As an example, if we assume that the dynamic range of a particular viability assay spans 0 to 125,000 cells per dish well, cell densities above 125,000 cells/well will not be resolved as a proportional change in the readout (**Figure 1A**). In this example, the investigator would not observe effects of treatment on viability at the highest cell densities, and would report a false negative.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2020 doi:10.20944/preprints202009.0481.v1

Figure 1

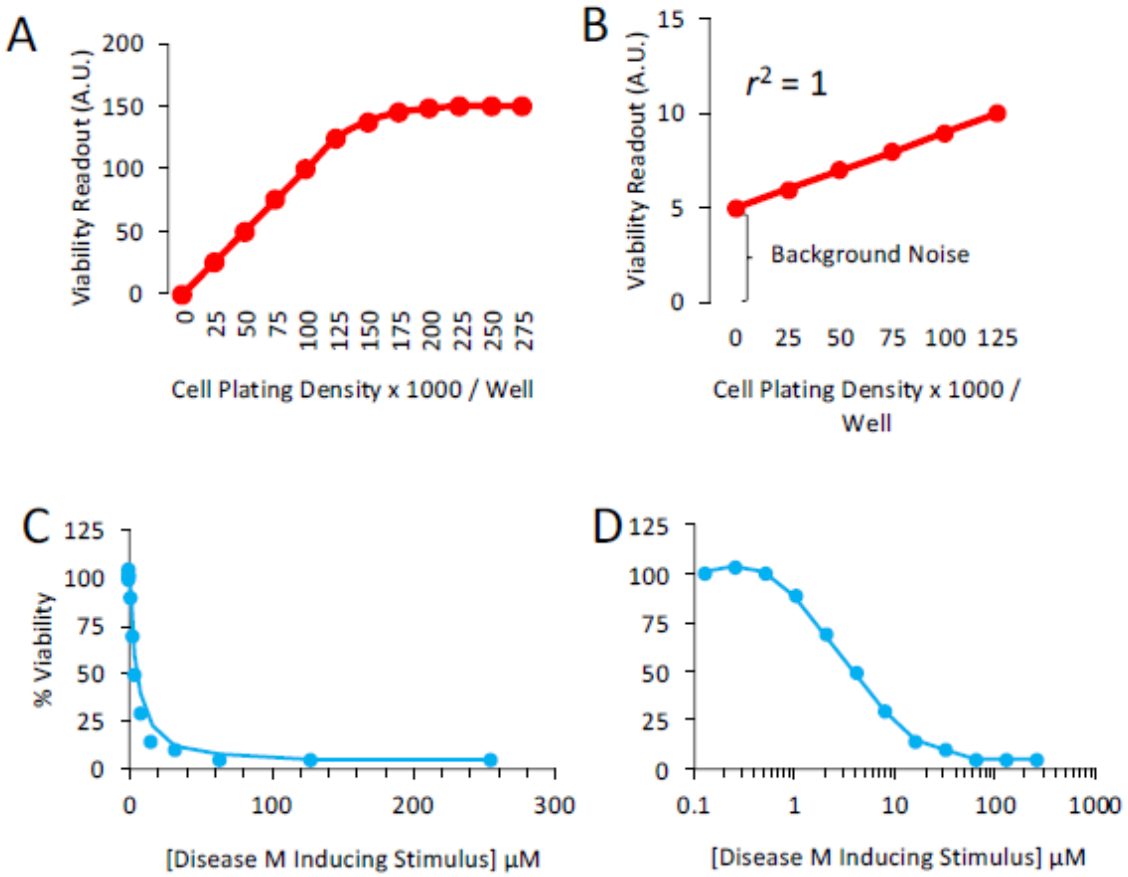

Figure 1. (A) Correlation between cell plating densities and cellular viability measurements. Note the measurement plateau at plating densities higher than 125,000 cells per well. The investigator will fail to observe any effects of treatment on viability at high cell densities (*i.e.*, a Type II error or false negative). (B) Perfect correlation between plating densities and viability measurements, but background noise accounts for a large fraction of the positive signal plotted on the Y-axis. High background noise can also mask the size of differences across groups. (C) Concentration-response viability curve for hypothetical disease M-inducing stimulus, plotted on a linear scale. (D) Concentration-response viability curve for hypothetical disease M-inducing stimulus, plotted on a logarithmic scale. The LC_{50} (or IC_{50}) in this example is 4 micromolar of the toxicity-inducing stimulus. Two-fold (rather than ten-fold) changes in concentration were employed in this example, to promote greater confidence in the LC_{50} measurement.

In **Figure 1B**, there is a perfect correlation between cell density and the viability readout, because the idealized data points converge upon a straight line. In real-life examples, the scatter of individual data points is higher than shown in **Figures 1A-B**, which reduces the coefficient of determination (abbreviated r^2), defined as the proportion of variance in the dependent variable (the outcome measurement or assay readout, plotted on Y axis) that is predictable from the independent variable (cell density, plotted on X axis). Although the data are in **Figure 1B** are linear and the coefficient of determination lies at the maximum of 1.0, the viability assay is too insensitive. The slope of the red line is low, and the background noise is high. As the density of the cells doubles from 25,000 to 50,000, the readout shifts from 6 to 7, a mere 17% change. *Ideally, the assay readout should increase by twofold when cell numbers double*. Background values could be subtracted from the other readings in this example, because the readout in empty wells cannot be attributed to cellular material and, therefore, do not reflect viability. On the other hand, when background noise is high, it is better to identify and remove its source, such as cell culture media-coloring agents (25), or to find an alternative assay. In

 doi:10.20944/preprints202009.0481.v1

Figure 1A, the Y-intercept is zero when there are no cells plated in the wells, suggesting no background noise.

Undervaluing the Temporal Dimension

The importance of timing is often underestimated in preclinical research, because therapeutic interventions may only be effective in patients before irreparable damage (and irreversible cell loss) has taken root. On the other hand, if the intervention is tested before *any* clinical symptoms have surfaced, the translatability of the study is compromised, as patients seek treatment only once a recognizable—albeit perhaps mild—syndrome has emerged. Ideally, the optimal temporal windows for drug delivery and the temporal kinetics of biological responses should both be identified; if drug-induced protection is only measured at acute stages but wanes over time, then the therapeutic intervention is less clinically meaningful. Merely transient protection would not benefit the patient in the long run. For *in vivo* functional measurements after acute brain injuries, waiting for at least one month post-intervention is recommended, to ensure that the therapeutic effects are not fleeting (26).

Aside from paying attention to the temporal progression of the disease and the temporal characteristics of drug delivery (time of initiation of treatment, frequency of delivery, circadian time of drug delivery (27), *etc.*), the amplitude (intensity) of the disease and the dose of the treatment also need to be titrated (see below), provided the lab can afford the increased cost. For example, researchers need to avoid models with intense pathology that cannot be tempered by *any* form of intervention, or else the hypothesis that the drug is protective is untestable. In sum, the timing and intensity of disease induction and therapeutic interventions should be chosen with an eye toward clinical translation.

Titrating the Dose

An important outcome of disease modeling includes the loss of cells from a particular organ. If the disease model is severely toxic and the cell loss is massive, it will be difficult or impossible to find an effective intervention (see above). For these reasons, it is common to titrate the intensity of the disease model to avoid overwhelming toxicity or disease pathology, such that cell viability approximates 50% of the control values observed in uninjured cells or animals. An example of viability data designed to explore the toxicity of a disease model is displayed using a linear scale in **Figure 1C** and a logarithmic scale in **Figure 1D**. Two-fold changes in dose (for *in vivo* experiments) or concentration (for *in vitro* experiments) of the toxicity-inducing stimulus were applied in this example. Although ten-fold changes in concentration/dose are more common, two-fold changes promote greater confidence in the LD_{50} or LC_{50} estimate, defined as the lethal dose or concentration leading to loss of ~50% of the biological population under study, respectively. With transgenic disease models, the two-fold titration approach described above is not feasible. Instead, one can test the intervention at early disease stages, before the syndrome becomes too severe.

In order to find the most effective dose of hypothetical drug "D", a foreshortened dose/concentration-response curve for the candidate drug is often plotted (**Figure 2A**), while clamping the dose/concentration or intensity of disease model "M" at LD_{50} or LC_{50} values. Full dose-response curves are not commonly performed in preclinical animal research, for the sake of economy, and the choice of dose is sometimes based only on previous literature or tradition. It is more common to test a wide range of concentrations in cell culture studies, but the effective concentrations in cell culture media do not readily translate to the plasma or interstitial fluid of the animal.

In **Figure 2A**, the disease model M leads to 50% loss of cellular viability in the first set of bars, and drug D prevents this toxic effect in a concentration-dependent manner, in the second to fourth set of bars. Not all drugs will display 100% efficacy in mitigating the sequelae of disease M, as shown in the fourth set of bars of this idealized figure. As stated above, the graph is incomplete, because virtually all drugs—including those designated as therapies—will reduce viability when applied in sufficiently high quantities (**Figure 2B**), leading to an inverted U-shaped pattern, rather than the sigmoidal curves often illustrated in textbooks. In short, the dose of the candidate drug will determine

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2020 doi:10.20944/preprints202009.0481.v1

whether or not it has poisonous qualities, as affirmed centuries ago by the toxicologist von Hohenheim (commonly known as Paracelsus) (28).

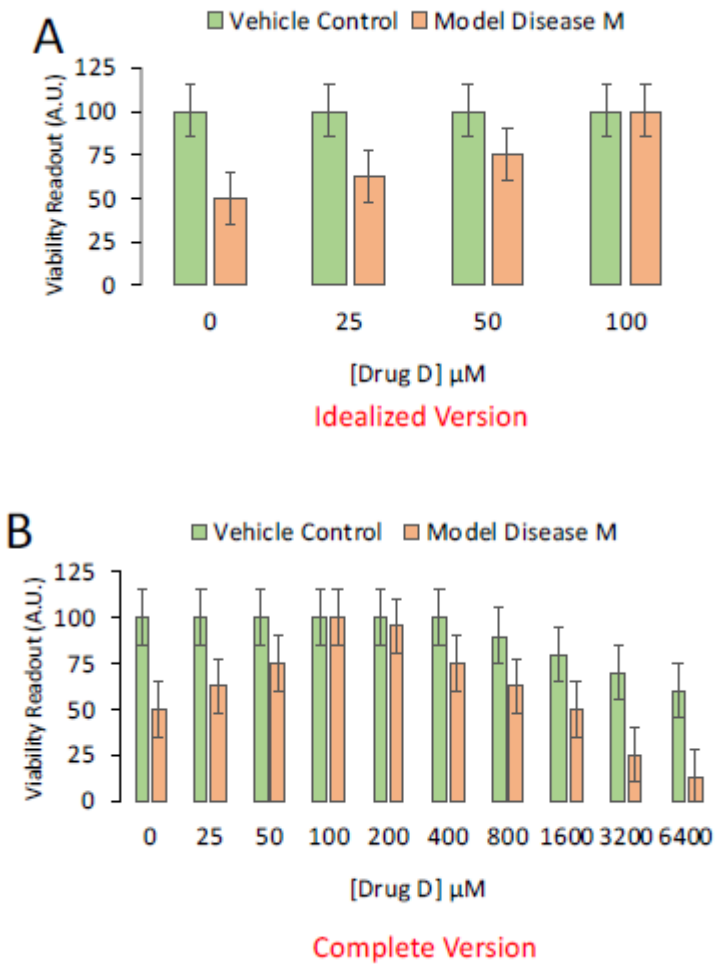

Figure 2. Idealized versus complete versions of concentration-response bar graphs for hypothetical drug D. (A) Drug D alleviates the toxicity of disease model M at rather high concentrations, ranging from 25-100 µM. (B) A complete concentration response graph reveals that slightly higher concentrations of drug D are toxic to baseline viability and exacerbate the toxicity of disease model M. The drug in this example therefore has a low therapeutic index, defined as the ratio of the concentration that elicits off-target toxicity to the concentration that elicits the desired therapeutic effect.

Choosing the Appropriate Controls

The inclusion of positive control groups helps to ensure that the assay can indeed resolve the expected differences. Conversely, negative control groups serve to confirm that the assay does not report differences when none are supposed to exist. In our example, drug D was applied in the presence of both the disease M-inducing toxicant as well as in the presence of vehicle (the negative control). Often, the *in vitro* vehicle of choice is dimethyl sulfoxide, due to its miscible properties. As stated above, 'the dose makes the poison', and dimethyl sulfoxide can be toxic above a surprisingly low threshold (29). It is, therefore, sometimes necessary to include an "untreated" control, in addition to applying multiple concentrations of dimethyl sulfoxide. Alternatively, all experimental groups should receive the same final volume and concentration of vehicle. If the group treated with the highest concentration of a drug/toxicant also receives the greatest volume of vehicle, but the controls fail to account for this, the researcher might erroneously conclude that the protective or toxic effects

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2020 doi:10.20944/preprints202009.0481.v1

are due to drug D or the disease-inducing toxicant, when they are actually attributable to a higher volume of vehicle than in the "control" group.

The importance of administering drug D both in the absence and presence of the stimulus that elicits the model disease M is further displayed in **Figure 3A**. If drug D increases cell viability under baseline conditions, even in injured cells, and it increases cell viability to the same degree under disease conditions, then the conclusion that "drug D changes the impact of disease model M" is false. This conundrum is addressed by choosing the proper test, a two-way ANOVA (on Gaussian and homoscedastic data), which assesses the impact of two independent variables (factors) on a dependent variable (outcome measurements), as well as any statistical interaction between the two independent variables.

Figure 3

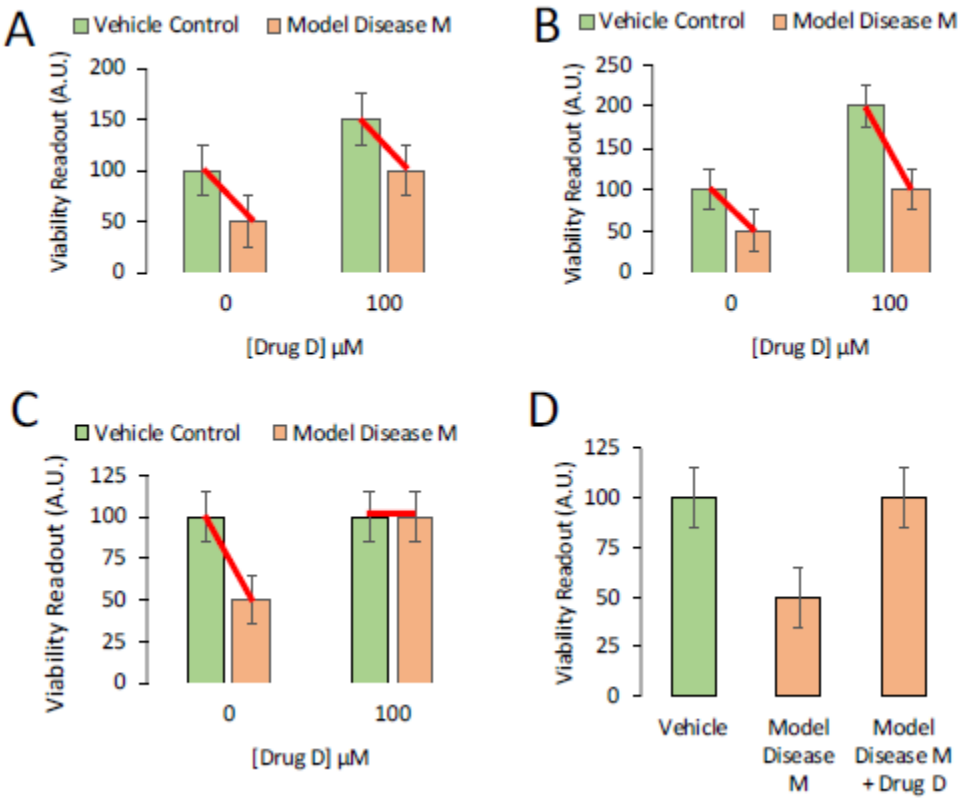

Figure 3. Impact of drug D on cellular viability in disease model M. (A-B) Drug D raises baseline viability but does not truly protect cells against the loss of viability with disease model M. (C) Drug D completely prevents the loss of cellular viability with disease model D at the indicated concentration. (D) If one of the control groups (drug D administered in the absence of disease; second green bar from panel B) is not included in the study, the data might appear as shown in this panel, but the investigator would not be aware that drug D simply increases *basal* viability and fails to modify the toxic impact of disease model M. In other words, the graph in panel D is similar to the graph in panel B, but is missing a control group.

In this example, the first factor is "disease state," of which there are two levels: i) administration of saline or ii) delivery of the injurious stimulus that induces model disease M at LC_{50} values. This factor is plotted as green *versus* orange bars in **Figure 3**. The second factor is "treatment," of which there are also two levels: i) vehicle or ii) drug D. For simplicity, only one concentration of drug D is displayed in **Figure 3** and plotted on the X axis. Multiple concentrations of drug D could be used in this example, but additional concentrations might raise the risk of a false negative or Type II error.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2020 doi:10.20944/preprints202009.0481.v1

Factors should not be confused with levels. If more concentrations of drug D or disease stimulus M were added, a two-way ANOVA would nevertheless be employed, as the influence of only two factors—disease state and treatment—is still being tested.

The dependent variable in **Figure 3** is the preclinical outcome (*e.g.*, cellular viability). In **Figure 3A**, there is no statistical interaction between drug D and disease model M. Drug D improves cellular viability by the same degree, *whether or not the cells are diseased*. When the data are graphed as red lines connecting the means, they lie in parallel, because the two independent variables do not influence each other's impact on viability (**Figure 3A**). In **Figure 3B**, the red lines are not parallel, and there is a statistical interaction between the two independent variables. However, in neither **Figure 3A** nor **3B** can the investigator conclude that drug D is protective against disease model M. Even though there is an interaction between the two variables in **Figure 3B**, the loss of viability with disease model M is still 50% of control, non-diseased values, in both the absence and presence of drug D.

Figure 3C reveals a protective effect of drug D and an interaction between the two variables. Even if the protective efficacy of drug D was not as high as in idealized **Figure 3C**, but reproducible, the researcher might conclude that drug D shows promise in the battle against preclinical disease M.

A word of caution—a simple pairwise comparison test after the ANOVA will not account for the change in baseline viability in the non-diseased group. Rather, the *post hoc* test might reveal a difference between the two groups shown in orange in **Figures 3A-B**, whether or not drug D is truly protective. Furthermore, students may not include the non-diseased control group treated with drug D (second green bar, **Figure 3A-C**) and perform a one-way ANOVA on three groups only, as in **Figure 3D**. The latter experimental design is sometimes justified, but it will fail to reveal if the increase in viability with drug D is due to an increase in *basal* viability, or is attributable to actual modification of disease processes (*i.e.*, there is a statistical interaction between disease and treatment factors).

Identifying a Suitable Biological Target

When testing a new intervention, there is usually not a vacuum of knowledge about its biological effects. The simplest way to commence the project is to read the literature and determine if there exist any data linking the candidate drug, or similar classes of drugs, with hypothetical target P or proteins that are homologous to P. Even if drug D was never tested in the brain, for example, it may have been studied in models of breast cancer. This knowledge can then be leveraged in brain tissue or cells. If the candidate drug is entirely new, RNA sequencing or proteomic analysis is often used to identify the biological pathway that is engaged by the drug.

While characterizing the impact of drug D on protein P, the timing of the assay is critical (see section on Undervaluing the Temporal Dimension). If receptor activation, protein cleavage, or protein phosphorylation is hypothesized as the mechanism of action, to name a few, measurements must be made close in time to application of drug D. If changes in gene expression are expected, the cells must be allowed sufficient time to engage the transcription and translation machinery. Either way, the protein will usually need to be assessed at multiple time points. Furthermore, a vehicle control may need to be included for every time point, if the vehicle exerts time-dependent effects of its own. Students may only include a "zero hours" time point as a substitute for a vehicle control, but this will not capture potential time-dependent changes induced by the vehicle. For example, vehicle administration *in vivo* may activate the stress axis in animals, and the investigator would ideally need to control for this variable, in a manner that a "0 hour" control time point cannot.

Measurements of protein levels do not ineluctably reflect function, and it is therefore edifying to include both expression and activity assays. While searching for a candidate protein P affected by drug D, it is important to distinguish between protein "induction" and "activation", because proteins can be activated without induction and *vice versa*. Protein induction is examined by measuring levels of the mRNA. However, cells might increase mRNA levels without any change in protein expression—if the protein is also degraded more rapidly. Thus, changes in mRNA levels do not necessarily lead to parallel changes in protein concentrations, and, for this reason, it is insufficient to show that mRNA levels are affected by drug D.

For the sake of argument, the *function* of protein P is hypothesized to upregulated or downregulated by drug D. When the overall expression of a protein is increased, there is no guarantee that it is also more active or that its function is enhanced. As an example, drug D might promote dephosphorylation-mediated inactivation of protein P, which may subsequently result in a compensatory increase in total protein P levels, but without any *net* increase in the numbers of functional molecules of P. If one was unaware of the compensatory nature of the delayed increase in expression of total protein P, one might assume that drug D increases the function of protein P and pathways downstream of P, when the opposite is true. To control for these caveats, it would be ideal to include a robust and specific assay for functional protein P and to assess its activity across time, if the laboratory enjoys those resources.

Conclusion

In Part I of this series, educators and students have learned to ensure that measurement outcomes lie within the dynamic range of the technical assay and how to avoid some of the most common pitfalls in preclinical research. It is important to have forehand knowledge of such issues, to avoid false interpretations and unproductive research paths. In Part II of this series, the reader will learn how to test the hypothesis that a particular molecular target of drug D mediates its protective effects, using a readout with satisfactory sensitivity and avoiding some of the errors common in data interpretations.

Abbreviations

None

Authors' contributions: RKL conceived of and wrote the paper in response to graduate student training needs.

Funding: This work was supported by National Institutes of Health, R15 grant1R15NS093539 and by the School of Pharmacy at Duquesne University.

Acknowledgments: The author wishes to thank the Graduate School of Pharmaceutical Sciences at Duquesne University and the National Institutes of Health.

References

1. Bechtel W. The Downs and Ups of Mechanistic Research: Circadian Rhythm Research as an Exemplar. Erkenntnis. 2010;73(3):313-28.
2. van Mil MHW, Postma PA, Boerwinkel DJ, Klaassen K, Waarlo AJ. Molecular Mechanistic Reasoning: Toward Bridging the Gap Between the Molecular and Cellular Levels in Life Science Education. Science Education. 2016;100(3):517-85.
3. Greenland S. Valid P-Values Behave Exactly as They Should: Some Misleading Criticisms of P-Values and Their Resolution With S-Values. The American Statistician. 2019;73(sup1):106-14.
4. Casadevall A, Fang FC. Descriptive science. Infect Immun. 2008;76(9):3835-6.
5. Marincola FM. In support of descriptive studies; relevance to translational research. J Transl Med. 2007;5:21.
6. Kell DB, Oliver SG. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. Bioessays. 2004;26(1):99-105.
7. Schreiber JB. New paradigms for considering statistical significance: A way forward for health services research journals, their authors, and their readership. Research in Social and Administrative Pharmacy. 2020;16(4):591-4.
8. Ioannidis JPA. What Have We (Not) Learnt from Millions of Scientific Papers with P Values? The American Statistician. 2019;73(sup1):20-5.
9. Anderson AA. Assessing Statistical Results: Magnitude, Precision, and Model Uncertainty. The American Statistician. 2019;73(sup1):118-21.
10. Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol. 2016;31(4):337-50.
11. Curran-Everett D. Evolution in statistics: P values, statistical significance, kayaks, and walking trees. Adv Physiol Educ. 2020;44(2):221-4.
12. de Leon J. Teaching Medical Students How To Think: Narrative, Mechanistic and Mathematical Thinking. Actas Esp Psiquiatr. 2018;46(4):133-45.

13. Bryant CD, Zhang NN, Sokoloff G, Fanselow MS, Ennes HS, Palmer AA, et al. Behavioral differences among C57BL/6 substrains: implications for transgenic and knockout studies. J Neurogenet. 2008;22(4):315-31.
14. Zhao L, Mulligan MK, Nowak TS, Jr. Substrain- and sex-dependent differences in stroke vulnerability in C57BL/6 mice. J Cereb Blood Flow Metab. 2019;39(3):426-38.
15. Boleij H, Salomons AR, van Sprundel M, Arndt SS, Ohl F. Not all mice are equal: welfare implications of behavioural habituation profiles in four 129 mouse substrains. PLoS One. 2012;7(8):e42544.
16. Voelkl B, Vogt L, Sena ES, Wurbel H. Reproducibility of preclinical animal research improves with heterogeneity of study samples. PLoS Biol. 2018;16(2):e2003693.
17. Wurbel H. Behaviour and the standardization fallacy. Nat Genet. 2000;26(3):263.
18. Tran J, Anastacio H, Bardy C. Genetic predispositions of Parkinson's disease revealed in patient-derived brain cells. NPJ Parkinsons Dis. 2020;6:8.
19. Chen SD, Li HQ, Cui M, Dong Q, Yu JT. Pluripotent stem cells for neurodegenerative disease modeling: an expert view on their value to drug discovery. Expert Opin Drug Discov. 2020:1-14.
20. Posimo JM, Unnithan AS, Gleixner AM, Choi HJ, Jiang Y, Pulugulla SH, et al. Viability assays for cells in culture. Journal of visualized experiments : JoVE. 2014;83(83):e50645.
21. Cheng HC, Ulane CM, Burke RE. Clinical progression in Parkinson disease and the neurobiology of axons. Ann Neurol. 2010;67(6):715-25.
22. Andersen PER, LØMo T. MODE OF CONTROL OF HIPPOCAMPAL PYRAMIDAL CELL DISCHARGES. In: Whalen RE, Thompson RF, Verzeano M, Weinberger NM, editors. The Neural Control of Behavior: Academic Press; 1970. p. 3-26.
23. Sorge RE, Martin LJ, Isbester KA, Sotocinal SG, Rosen S, Tuttle AH, et al. Olfactory exposure to males, including men, causes stress and related analgesia in rodents. Nat Methods. 2014;11(6):629-32.
24. van Bodegom M, Homberg JR, Henckens M. Modulation of the Hypothalamic-Pituitary-Adrenal Axis by Early Life Stress Exposure. Front Cell Neurosci. 2017;11:87.
25. Ettinger A, Wittmann T. Fluorescence live cell imaging. Methods Cell Biol. 2014;123:77-94.
26. Lapchak PA, Zhang JH, Noble-Haeusslein LJ. RIGOR guidelines: escalating STAIR and STEPS for effective translational research. Transl Stroke Res. 2013;4(3):279-85.
27. Esposito E, Li W, E TM, Park JH, Sencan I, Guo S, et al. Potential circadian effects on translational failure for neuroprotection. Nature. 2020.
28. Leak RK, Calabrese EJ, Kozumbo WJ, Gidday JM, Johnson TE, Mitchell JR, et al. Enhancing and Extending Biological Performance and Resilience. Dose Response. 2018;16(3):1559325818784501.
29. Galvao J, Davis B, Tilley M, Normando E, Duchen MR, Cordeiro MF. Unexpected low-dose toxicity of the universal solvent DMSO. FASEB J. 2014;28(3):1317-30.