Modeling experimental error in assays: Understanding discrepancies between assay results with different dispensing technologies

Sonya M. Hanson,[1] Sean Ekins,[2] and John D. Chodera[1,*]

[1]*Computational Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065, United States*
[2]*Collaborations in Chemistry, Fuquay-Varina, NC 27526, United States*

(Dated: August 19, 2015)

All experimental assay data is contaminated with error, but understanding the magnitude, type, and primary origin of this error is not often obvious. Here, we describe a simple set of assay modeling techniques that allow sources of error and bias to be simulated and propagated into assay results. We demonstrate how deceptively simple operations—such as the creation of a dilution series with a robotic liquid handler—can significantly amplify imprecision and even contribute substantially to bias. To illustrate these techniques, we review an hour example of how choice of dispensing technology can greatly impact assay measurements, and show how the primary contributions to discrepancies between assay results can be easily understood. These simple modeling techniques—illustrated with an accompanying IPython notebook—can allow modelers making use of experimental data to understand the expected error and bias in the dataset, and even help experimentalists during the assay design stage to ensure that assays are capable of reaching their target accuracy and imprecision goals.

I. INTRODUCTION

Measuring the activity and potency of ligands—whether in biophysical or cell-based assays—is a critical step in optimizing small molecules for use as chemical probes or potential therapeutics, and indeed in probing biological processes in general. Use of this assay data is complicated by the fact that all assay data are contaminated with error, with contributions to this error arising from numerous sources.

Often, the dominant contributions to assay error are simply not known—this is unsurprising, given the number and variety of potential contributions to assay error. Even for what might be considered a straightforward assay involving fluorescent measurements of ligand binding to a protein target, this might include (but is by no means limited to): compound impurities and degradation, imprecise compound dispensing, unmonitored water absorption by DMSO stocks, intrinsic compound fluorescence, compound insolubility or aggregation, variability in protein concentration or quality, pipetting errors, and inherent noise in any fluorescence measurement. In an ideal world, a number of control experiments would be run to measure the magnitude of these effects, and data quality checks would either reject flawed data or ensure that all contributions to error have been carefully accounted for in producing an assessment of error and confidence for each assayed value.

Unfortunately, by the time the data reach the hands of a modeler (or other data consumer), the opportunity to perform these careful control experiments has passed, and yet somehow, one is expected to make good use of the data. In the worst case, the communicated assay data may not contain any estimate of error whatsoever, making it fiendishly difficult to draw conclusions from the data—is the difference between multiple assay values for closely related compounds due to a true structure-activity relationship driving potency, or simply due to random error? While aggregating many datasets can give a crude estimate of the general reliability of similar assays [1, 2], knowledge of how a particular assay was conducted can inform the construction of an assay-specific model incorporating some of the dominant contributions to error in a manner that can be suprisingly informative.

In this paper, we review some common sources of error in experimental assays, and describe some simple modeling tools for simulating a model of an assay that incorporates these important (often dominant) sources of error. This approach, while simple, should nevertheless provide a powerful tool for modelers to understand how assay error depends on important parameters, such as providing a means to estimate the expected data error as a function of compound affinity. Not only modelers may find benefit in this approach—these tools can be also used to help optimize assay formats before an experiment is performed, help troubleshoot problematic assays after the fact, or ensure that all major sources of error accounted for by checking that variations among controls match expectations.

We illustrate these concepts by considering a now-infamous example from the literature: a report by Ekins et al. [**?**] on how the choice of dispensing technology impacts the apparent biological activity of the same set of compounds under otherwise identical conditions. The datasets employed in the analyses [CITE] used either a standard liquid handler with fixed (washable) tips or an acoustic droplet dispensing device to prepare the assay, resulting in highly divergent assay results (Figure **??**). While the frustration for modelers was particularly great, since QSAR models derived from these otherwise identical assays produce surprisingly divergent predictions, numerous practitioners from all corners of drug discovery expressed their frustration in ensuing blog posts and commentaries [**?**]. Hosts of potential reasons were speculated [**?**], including [LIST HERE].

We make use this real-world example to illustrate some

* Corresponding author; john.chodera@choderalab.org

of the basic concepts behind modeling the most common sources of error in assays. For simplicity, we ask whether the most basic contributions to assay error—imprecision and bias in material transfer operations and imprecision in measurement—might account for some component of this discrepancy between assay techniques. We make use of very basic information—the assay protocol as described (with some additional inferences based on basic concepts such as compound solubility limits) and manufacturer specifications for imprecision and bias—to construct a model of each dispensing process to determine the overall inaccuracy and imprecision of the assay as a function of true compound affinity, and identify the steps that contribute the most to these assay data errors. To better illustrate these techniques, we also provide an annotated IPython notebook that includes all of the computations described here in detail. Readers are encouraged to download these notebooks and play with them to see how different assay configurations affect assay error.

II. EXPERIMENTAL ERROR

Overall experimental error can be broken into two components: The imprecision (or variance), which characterizes the random component of the error that causes different replicates of the same assay to give slightly different results, and the inaccuracy (or bias), which is the deviation of the average over many replicates from the true value of the quantity being measured.

Serial Dilution

More words.

How off?

This off.

III. CONCLUSION

Everyone should do things better.

IV. ACKNOWLEDGMENTS

Thanks to everybody!

[1] C. Kramer, T. Kalliokoski, P. Gedeck, and A. Vulpetti, J. Med. Chem. **55**, 5165 (2012).

[2] T. Kalliokoski, C. Kramer, A. Vulpetti, and P. Gedeck, PLoS One **8**, e61007 (2013).