

Pen before pipette

Researchers should minimize arbitrary choices in experimental design. Some tools are available to help.

The ancient Greeks personified forethought and afterthought as the brothers Prometheus and Epimetheus, the only two Titans spared from exile in the human creation myth. As in most myths, the gods amplified human tendencies, epitomizing the dichotomy between these two ways of thinking. Modern research needs a Promethean adjustment. The growing scale of data production demands a renewed emphasis on planning and design before hitting 'play' on an experiment. At stake are missed opportunities and potentially worse—misleading results and wasted time, resources or animals.

Experimental design is the practice of planning a study to reach specified objectives. As a formal subject, it is taught in the context of statistics, as an approach to constructing experiments that allow a meaningful hypothesis test. The process forces a researcher to frame the question clearly and to explicitly define variables and assumptions. Formal experimental design focuses on the designation of appropriate controls, random assignment of subjects that receive an experimental treatment, replication, and blocking to remove the potential effect of confounding variables. More broadly, it encompasses myriad experimental decisions such as how many samples to use; what types of treatments, concentrations and time points are appropriate; which experimental apparatus to employ; what exactly to measure; and how to analyze data.

Poor study design can lead to underpowered, biased or irreproducible results. In an effort to raise awareness and improve transparency, the Nature research journals have asked since 2013 that life-science authors include a [reporting summary](#) during review that lists key experimental and analytical design details in a single simple document. Since mid-2017, this reporting summary has been published with the paper. But by the time a paper is submitted and reviewed, options for addressing design flaws are limited.

Formal design is commonly practiced in research on human subjects, and more generally, grant applications are required to include detailed research plans. However, experimental choices in biology are not always justified or arrived at in a systematic way. Researchers should consider taking advantage of methods developed to introduce more rigor into experimental planning.

In a Correspondence published online (<http://dx.doi.org/10.1038/nmeth.4462>), authors from the National Centre for the Replacement, Refinement and Reduction of Animals in Research (NC3Rs) describe a free web tool to help researchers improve their experimental designs. Their interactive Experimental Design Assistant (EDA)

allows experimental aspects from objectives to analysis to be visualized and examined. The tool checks for missing elements and inconsistencies with an iterative critiquing function, and supports blinding and power calculations to determine the sample size needed to find statistically significant results.

Other tools exist to help plan experiments at the cellular scale. In this issue, Altschuler, Wu and colleagues (p967) present sampling strategies to accurately capture population heterogeneity from single-cell studies using as few cells as possible. Based on pilot data, the approach can replace heuristic decision-making in applications such as tumor microarray analysis and well-based high-content screening. For example, the authors show that the rule of thumb of using three tissue cores in tissue microarrays is only sufficient to capture heterogeneity for a small fraction of cellular image features from precious tumor samples. Most features require more cores.

Strategies also exist for experiments at the molecular level. For the routine question of how deeply to sequence a sample, a mathematically rigorous approach can predict the molecular complexity, or fraction of unique molecules in the population, that a given level of sequencing effort will attain (T. Daley & A.D. Smith *Nat. Methods* **10**, 325–327; 2013). This replaces the need for intuition and avoids both undersequencing, which can lead to lost information or a biased representation of the sample, and oversequencing, which wastes resources and accumulates noise due to sequencing errors.

Quantitative predictions that underlie good experimental design rely on pilot data. In general, design considerations are less important for purely exploratory work. But preliminary knowledge can be used to identify the optimal way to conduct subsequent hypothesis-driven research. Also in this issue, the organizers of the DREAM challenges for computational tools argue (p937) that computational models (typically based on existing data) should be used to guide data collection in systems biology.

Educating young scientists is an essential part of a renewed focus on experimental design. The US National Institutes of Health, for instance, provide free online educational material as part of a [training initiative](#). Such resources should be incorporated into basic scientific education.

Greater emphasis on systematic decision making in experimental design will improve rigor in biological research. In the Greek telling, Prometheus, the bringer of fire, was the better friend to humankind; Epimetheus opened Pandora's box.