

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Introductory Chapter: Research Methods for the Next 60 Years of Space Exploration

Robert J. Reynolds and Mark Shelhamer

1. Introduction

There are many potential health hazards inherent to space travel, and, as the chapters in this book make clear, even after 60 years of human space exploration, much is left to be learned about how to live and work in space. As a result of the diversity of problems that remain to be solved, the scientific methods required to research these issues need to be flexible and varied. This is perhaps most true in our approach to analyzing data and drawing conclusions from them in the context of space medicine.

In a commentary published in the *Journal of Applied Physiology*, Ploutz-Snyder et al. [1] point out that in the study of exotic topics (such as the physiology and health of space travelers) the available data are often insufficient to satisfy the sample-size requirements for traditional null-hypothesis statistical testing (NHST). They rightly point out that if we hold this as the standard of good research, (i.e., if NHST is our only, or even our preferred, tool for learning from data) we will be forced to abandon whole lines of research. While the authors offer several “approaches for justifying small-n research,” even these are attempts to shoehorn small datasets into traditional statistical analysis. This misses the broader (epistemological) point: what is needed in small-n studies is not just a better way to use statistics, but rather other tools which afford the freedom to learn without using statistics at all.

2. The problems of small-n settings

Research on small sample sizes poses a number of challenges. First and foremost is the violation of assumptions that frequentist statistical methods often require in order to be valid. Secondary to this, but inherent in the nature of small samples, is the typical lack of statistical power for detecting differences other than those in low-variance settings or those with dramatic effects. Each of these two challenges can lead to difficulty in interpreting results.

2.1 Violations of frequentist assumptions

Most frequentist statistical analyses follow a familiar pattern: assume the outcome follows a known statistical distribution, then test whether or not the observed data are unusual (unexpected) under the null hypothesis. However, beyond basic goodness-of-fit considerations, such analyses require other assumptions as well, many of which are clearly violated much of the time. Perhaps the most important of these assumptions is that the observations in a given sample are “iid”—independent and identically

distributed. When samples are strictly observational (i.e., not from a randomized trial) this assumption is often unwarranted. The implication of violating this assumption can be profound: differential probability of exposure and inequitable distributions of potential confounders can lead to what is known as *confounding by indication*, a subtle form of bias that can lead to misleading or even wholly wrong conclusions.

2.2 Statistical power

In any statistical analysis our strength of conviction for our conclusions is largely dependent on how much data we can observe (sample size), and how consistent our outcomes are within those observed data (variance). In the frequentist statistical context this is reflected in the concept of *statistical power*. Statistical power is defined as the (hypothetical) probability of correctly rejecting the null hypothesis when the null hypothesis is indeed false (and false by a pre-set threshold considered to be of clinical or practical importance). A commonly desired and accepted level of statistical power is 80%. However, it should be noted that even with this level of power, there is a 20% chance of making a Type II error (i.e., incorrectly failing to reject the null hypothesis). Unless the ratio of the standard deviation to the mean (coefficient of variation) is small, the statistical power in small studies is considerably lower than 80%, effectively crippling the ability to confidently draw inferences from the data under this framework.

2.3 Interpretation

Both violation of assumptions and low statistical power can frustrate the drawing of inferences under traditional statistical approaches. If we manage to obtain a statistically significant effect, how should we interpret it given the potential for confounding by indication? If we fail to see any significant effects where we believe we ought to *a priori*, how do we interpret that? Does our assessment of the meaning of such results change with larger or smaller variance in our sample? Under traditional approaches we surrender to the probabilities of committing Type I or Type II errors, and resign ourselves to having learned nothing.

2.4 Preference for errors

The ultimate motive for use of the NHST framework is to reach reasonable conclusions about a population or process from a (large) subsample of it. However, a real yet unintended consequence of the framework is the focus on avoidance of error. The framework itself is centered on the concept of errors in inference: when we can, we design our studies to avoid Type I error while simultaneously trying to limit Type II errors. In so doing we may make these errors—rather than what we might learn from our data—the primary consideration of our scientific activity. It should come as no surprise that when we make avoidance of error our top priority, we fail to learn all we can from our data.

Modern science's focus on Type I error has proven to be particularly troublesome. In our quest to never actively assert a false truth we have no doubt passively allowed many truths to go unspoken. It is obvious that Type I errors can cause harm in medicine if new treatments are adopted that are actually harmful to patients. Less obvious is the harm that may result if research into a truly efficacious treatment is abandoned simply because a p-value was too high. Such harm is every bit as real (and every bit as irreversible) as that done by introducing an ineffective treatment. It is especially troubling in initial exploratory studies and those where data are acquired only with great difficulty or expense.

3. Methodological solutions for research in space medicine

Having seen the problems that small-n settings create in general, how do we solve them? Through a combination of realigning our epistemology, using our current tools differently, and utilizing modern analytic tools developed outside the field of statistics, we can do better research and advance the field of space medicine to meet the challenges of the next 60 years.

3.1 Realigning our epistemology

Cognitive dissonance is the feeling of discomfort one feels when actions fail to conform to beliefs. [2] To most scientists, making claims about truth without a statistically significant result to point to elicits substantial cognitive dissonance. This perhaps more than anything demonstrates our over-reliance on NHST as a substitute for a more robust epistemology. There are several things we can do to learn from data without suffering from cognitive dissonance—even without significance tests. Altogether they amount to a different epistemological approach to epidemiology for space exploration.

3.1.1 Guidelines for causation

In 1965 Sir Austin Bradford Hill described nine guidelines for determining causation from scientific evidence. [3] It is worth noting that while one of the guidelines deals with *strength of association*, or what we might recognize as *effect size*, none of the criteria deal with significance testing or p-values. Explicitly, Hill called for examining the *quality* of the relationship between exposure and outcome: the logical features of how the evidence suggests they interact, and how that fits with prior knowledge of the same or similar subject matter. This sort of prescription is well-suited to the small-n environment of space medicine.

3.1.2 Modern causal inference theory: assumptions

Similar to Hill's work, modern causal inference methods may also be of great use in space-health research. These methods have sought to mathematically formalize causation in order to make valid use of observational data for causal estimation and to avoid introducing biases in analyzing such data [4]. Perhaps more important than the methods of analysis that this framework has promoted is the understanding of the assumptions necessary to make causal statements from non-randomized data. Merely understanding the assumptions of positivity, consistency, and conditional exchangeability—and what happens when one violates them—can be of tremendous help when trying to draw inferences based on limited data.

3.1.3 Directed acyclic graphs

A common tool used in modern causal inference is a special type of network graph known as the directed acyclic graph (DAG). These are network maps that reflect causal relationships. DAGs are drawn according to some simple rules, but making and using these diagrams can be quite useful for clarifying thinking and formulating testable hypotheses. If we factorize a joint probability distribution over a DAG, we create a Bayesian Network, a powerful tool of probabilistic inference. If we decompose a correlation or covariance matrix over a DAG, we can do path analysis or structural equation modeling, forms of latent-variable analysis. Even without any data collected at all, the structure of a DAG implies variable dependencies and

independencies, which in turn have implications for what is and is not possible in the system from which the data were acquired, and thus can help guide critical thinking about problems.

3.1.4 Alternative hypotheses

A final epistemological realignment is to define specific, sensible hypotheses given the question at hand, which may or may not conform to the typical NHST two-tailed tests of significance. Examples of such alternatives include equivalence testing, inferiority testing, and a still more exotic choice, the *modus tollens*. All of these ask different questions than whether the central tendency of a sample shows enough difference to evince a significant p-value for the given sample size and variance. By changing the testable hypothesis to be more specific to what we really would like to know, we can often obtain an answer that is not only more sensible, but often more statistically powerful too, which might then bring NHST back into the realm of possibility to further refine the analysis.

3.2 Alternative analytic approaches

Yet another strategy for learning from data is the use of more-sophisticated analytic methods which do not necessarily rely on NHST. This includes exploiting properties of known statistical tests for alternative hypotheses, Bayesian methods, and machine learning.

3.2.1 Alternative uses of common statistical models

With a good understanding of common statistical models, it is possible to exploit their properties to conduct atypical investigations. Here we use an example from the literature on astronaut mortality to demonstrate this idea.

Using data on US astronauts and Soviet and Russian cosmonauts, Reynolds et al. [5] demonstrated that mortality from cancer and cardiovascular disease have no common causes in this population. This in turn was taken as evidence that doses of ionizing radiation received in space cannot have been sufficient to affect mortality from both of these causes. This was achieved by showing that a naïve analysis of survival curves (where competing causes of death were treated as censoring events) were not markedly different from survival curves that account for competing risks. That is, the causes of death displayed statistical independence which, in DAG terms, means they share no common ancestor.

In this example, the authors exploited the implications of different statistical methods for computing survival in presence of competing risks to make inferences regarding the structure of causal relationships. This is but one example, and undoubtedly others exist for those who can think broadly and conceptually about specific questions to be asked of existing datasets.

3.2.2 Simulation

The advancements in computing power over the last several decades have made possible more sophisticated forms of analysis, not least among them being simulation. We refer here to several different well-established approaches, all of which have found use in various domains such as statistics, business, and engineering.

Markov-chain Monte Carlo simulation (MCMC) has been used for decades in engineering for probabilistic risk assessment. Agent-based simulation has found increasing popularity in epidemiology for modeling community-level effects of policy

change or change in social environment. Techniques such as the bootstrap and the jackknife may be loosely grouped here as well, as they rely upon repeated recalculation of sample statistics using algorithms that resample the data in specific ways. Finally, simple “what-if” analyses can help find the extremes of what is possible in a process or phenomenon, and can be used to eliminate possibilities or competing hypotheses.

3.2.3 Bayesian methods

Though certainly not new, Bayesian methods are still underutilized in research in general and in space medicine in particular. This is primarily owing to the unfamiliarity of most researchers with these methods, which in turn is due to the lack of graduate-level training on them in most scientific programs other than statistics. Historically, this was sensible: their mathematical complexity and need for computing power made them difficult to implement for all but the simplest of applications. Fortunately, computer science and computer hardware have both evolved to where these methods are easy to implement, creating a large opportunity for researchers to work with smaller datasets in meaningful and rigorous ways without reliance on NHST and p-values.

3.2.4 Data science

In recent years, Data Science has been turning business analytics upside down. In general, data science is understood as the science of learning from data, a seemingly perfect fit to our objectives here. Yet Data Science has seen much slower adoption in Academia, perhaps owing to the fact that the only part of Data Science that fits with the traditional epistemological approach to research is that part of Data Science which uses traditional NHST statistics.

A hallmark of Data Science is the use of machine learning. However, many of the methods of machine learning are methods that typically benefit from large datasets: those with hundreds of columns and millions of rows. Nevertheless, machine learning does have techniques that can be of use in the small-n world. Techniques for data reduction, data visualization, data mining, and simulation all are powerful tools that can often be applied in the domain of small-n research. Perhaps of particular interest to space medicine, researchers are able to use these methods for exploratory data analysis and hypothesis generation, tasks at which unsupervised machine learning excels.

4. Summary and conclusions

In this chapter we have discussed the limits of NHST as a surrogate for a broader, more flexible epistemological framework. Over-reliance on NHST can cripple the research enterprise when sample sizes and sampling schemes fail to conform to the assumptions necessary for valid models, much less valid inference.

A motivating factor for the use of NHST is the desire to draw correct conclusions. This is a valid aim, but may lead to an emphasis on error avoidance at the expense of learning from (possibly limited) data. Instead, scientists need to consider evidence using Hill’s guidelines for causation, should examine whether or not the data in hand conform to or defy the assumptions needed for causal inference, and should include the use of DAGs to better understand what we already know about a given topic, and to clarify what we conjecture to be true *a priori*. Formulating so-called “alternative” hypotheses appropriate to the topic under study may even allow us to improve our inferences when using traditional NHST. There is no need to restrict ourselves to one approach or the other.

Alternative methods of analysis can be used to aid our understanding in small-data situations. Bayesian methods, more sophisticated uses of well-known statistical methods, and methods from data science all provide useful techniques that work well with small datasets, provided the scientist is willing to think differently about the outcomes of these analyses.

It is our hope that researchers involved in space medicine will adopt these perspectives and methods. To the extent that these ideas and techniques are adopted by the broader research community, we expect to see great advancements in our knowledge of health and safety in spaceflight. It is this expansion of our collective knowledge that will help make possible the space exploration missions of the next 60 years.

Author details

Robert J. Reynolds^{1,2*} and Mark Shelhamer³


1 Mortality Research and Consulting, Inc., USA

2 Translational Research Institute for Space Health, Baylor College of Medicine, USA

3 The Johns Hopkins University School of Medicine, USA

*Address all correspondence to: rreynolds@mortalityresearch.com

IntechOpen

© 2020 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Ploutz-Snyder RJ, Fiedler J, Feiveson AH. Last word on viewpoint: Justifying small-n research in scientifically amazing settings: Challenging the notion that only “big-n” studies are worthwhile. *Journal of Applied Physiology*. 2014;**116**(9):1254. DOI: 10.1152/japplphysiol.00214.2014
- [2] Festinger LA. *Theory of Cognitive Dissonance*. California: Stanford University Press; 1957
- [3] Hill AB. The environment and disease: Association or causation? *Proceedings of the Royal Society of Medicine*. 1965;**58**(5):295-300. DOI: 10.1177/003591576505800503
- [4] Pearl J. *Causality*. Cambridge: Cambridge University Press; 2009. DOI: 10.1017/CBO9780511803161
- [5] Reynolds RJ, Bukhtiyarov IV, Tikhonova GI, Day SM, Ushakov IB, et al. Contrapositive logic suggests space radiation not having a strong impact on mortality of US astronauts and Soviet and Russian cosmonauts. *Scientific Reports*. 2019;**9**:8583. DOI: 10.1038/s41598-019-44858-0