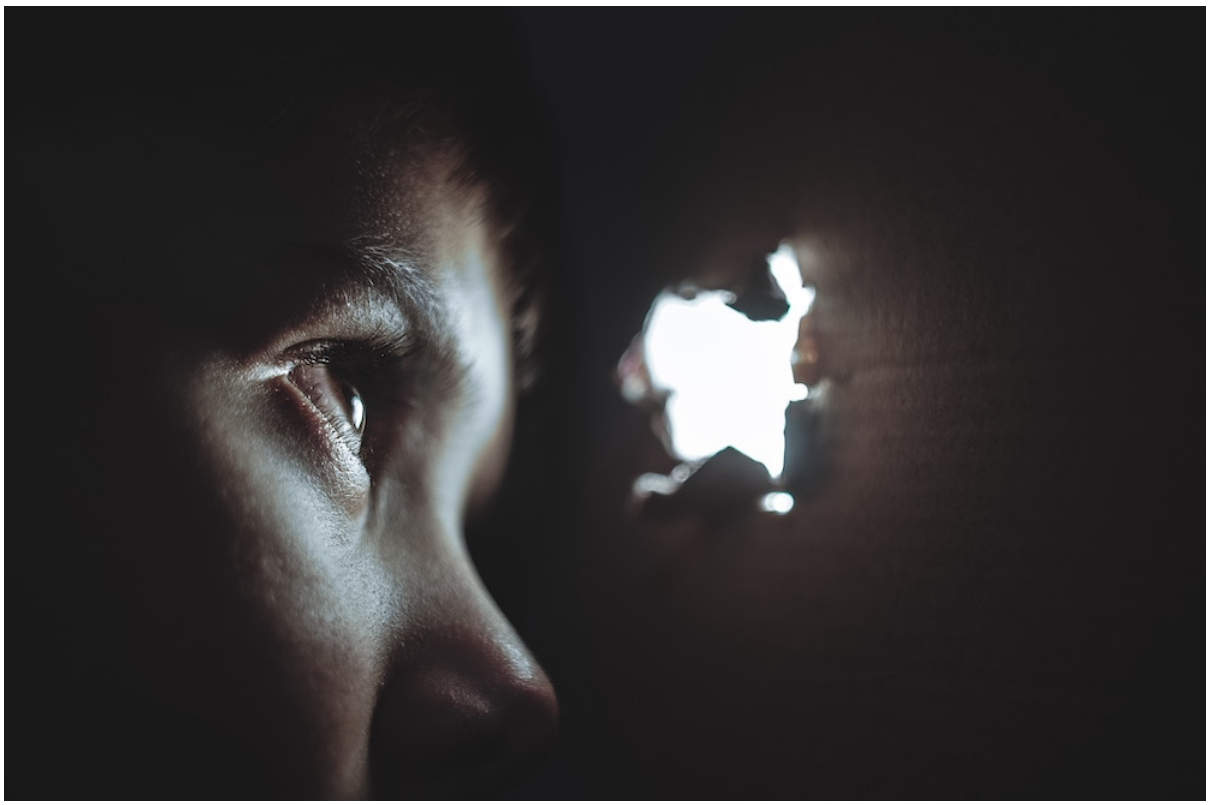
If Not Now, It's Too Late: More Clinical Science Pitfalls and a Path to Improvement

written by Robert McNutt, M.D. | April 25, 2019



Let’s review three major vulnerabilities with how randomized trials (RTs) are conducted, as discussed so far in this series. Critically appraising a research study involves determining the “internal and external” validity. Internal validity deals with the conduct of the study, per se. External validity deals with whether the study’s findings can be generalized to others in the population. Here’s what can go wrong:

> Populations being studied in RTs are too often convenience samples of patients/subjects rather than random or systematic samples of subjects. This diminishes our ability to externalize findings from the RT to the population at large. This is a defect in external validity.
> After a population of people is gathered for study, simple randomization often fails to identify enough groups of individuals with variable and prognostically important characteristics that may alter the assessment of the value

of a treatment to them, rather than to the average. This is a defect in both the external and internal validity of the study.
Simple randomization does not assure that there will be adequate balancing of prognostically important clinical and personal characteristics. An imbalance may falsely support or refute the independent value of the treatment being tested. This is a defect in internal validity.

Why Masking Matters

There are other issues that diminish the value of information from RTs for patients. First, a tenet of a RT is that researchers, and those researched, must *not know* what they are getting, and, also, those assessing the measured outcomes must *not know* who got what. This is called "masking," and it is a crucial aspect of how a RT is conducted.

There are legions of studies showing that researchers and subjects change assessments and actions if they know what they are getting. I am a wine taster. I may love a brand of wine, but if I am presented that wine without knowing the brand, I often change my assessment. A group of expert wine tasters were given the same wine in bottles with different names and prices; when they scored the wine, they favored some bottle names, and, also, their assessments of quality tracked with price—more expensive, higher the score.

This "bias by knowing" is noted in clinical research; examples are many. One of my favorites was a RT comparing spine surgery versus no surgery for a herniated disc. Those people who accepted the RT, and who got spine surgery, rated outcomes *the same* as those who did not get surgery. However, a companion

observational study found that people getting surgery outside the RT rated outcomes *better* for surgery than those who did not have surgery.

Even though all those who had surgery knew it in both the RT and the observational study, their expectations were different and so were their assessments. Outside the RT, participants knew what they were getting, and, hence, their judgments were biased. If participants know what they are getting, *randomization is useless.*

Longitudinal RTs Take Too Long

Next, RTs, as practiced today, *take way too long*. For example, The National Lung Cancer Screening Trial (NLST), discussed throughout this series, started enrolling patients in 2002, but the study was not published until 2011, nine years later. Think about this a minute. Suppose the study definitively showed that CT scan screening saved lives (it does not, in my view); then, many people were denied potential better care while the study ran its course.

Why the long time period? The outcome they measured was infrequent, and it took time to accrue. When the baseline likelihood of an outcome is small, large numbers of patients are needed to see if one plan is better than another. Some think the 50,000+ in this study is a large number of subjects, *but it is not;* it is an inadequate number *for this question.* If outcome event rates are small, we need larger samples of patients so outcomes can be known in less than a few years.

The Problems with Observational Studies and the Language of Research

In my initial blog of this series, I raised issue with observational studies and how we report trials. First, observational trials are dangerous; some may be helpful, but only when the outcome event rates and differences between compared groups are huge. This is rare, and from a practical perspective, all observational trials may be ignored from the perspective of informing patients who must make choices. If there are many vulnerable aspects of RTs, imagine how many there are with observational research.

Second, research studies are written in a language useful to only a small number of people; those who do research. However, researchers should not work for researchers; they should work for patients. Absolute differences in outcomes for benefit and harm are the only things that a patient should see. It is up to better research than we have today to assure the numbers are good enough.

A Roadmap for Better Clinical Results

In summary, RTs, as conducted today, fall short of a standard for informing people of the consequences of choices. On my own scale of what makes a RT valuable, generalizability is most important; paying particular attention to groups of people with variable clinical and personal characteristics that may affect the measurement of differences in outcomes is close as 1A. Addressing these two issues at the beginning of a RT may nullify the third concern (unequal numbers of people in prognostic subgroups). I am ignoring unmasking, as an unmasked RT should be ignored, as should observational comparative data.

The following list assumes that RTs will be looking for small differences in outcomes. Clearly, it takes fewer people to detect a difference in outcomes when the efficacy is large (don’t need a RT to know you should have a parachute if you jump from a plane).

So, to achieve better clinical research:

- Study random samples of patients from full populations. Or, Study an entire sample or patients.
- Stratify, and oversample people with the greatest variations in prognostic variables.
- If you don’t stratify, use pre-randomization schemes to assure balanced numbers of people in prognostic subgroups.
- Choose to measure a single or, at most, two outcome variables that can be measured accurately.

In addition, future studies must follow these principles:

- Be less expensive than they are today.
- Be able to constantly refresh insights.
- Produce results in contemporary time periods.
- Focus on community/local catchment areas.
- Use just a few standardized, disease-specific and prognostic subgroup data, passively obtained.
- Be transparent; use a “ticker-tape” or stock market approach for presentation of outcomes over time and a standardized table presenting absolute differences in disease/harm outcomes.
- Incentivize and pay patients to participate.
- Inform patients and develop ways to coach them for choice.

How to do this? Future blog posts will present alternatives.

Founded as ICLOPS in 2002, Roji Health Intelligence guides health care systems, providers and patients on the path to better health through Solutions that help providers improve their value and succeed in Risk. Roji Health Intelligence is a CMS Qualified Clinical Data Registry.

Image: Dmitry Ratushny