Increasing the Expected Value to Society of Clinical Research Studies?

Helena Chmura Kraemer, PhD

The value of a methodological paper generally depends on 1) the importance of the problem addressed, 2) the completeness of the presentation of the issues and prior progress, and 3) a new solution to the problem that is feasible and better than what has already been proposed. The Uyei and Braithwaite article1 (hereafter UB) clearly satisfies the first two criteria and is well worth contemplating. The article is thought-provoking, not only because it raises issues related to appropriate choice of control groups in randomized clinical trials (RCTs) on which many (I for one) have become complacent, but because it focuses attention on the societal impact of RCTs so often ignored. My focus here is on 3 issues: value to society, nonplacebo control groups, and benefit-to-harm balance, all central to UB arguments.

Nowhere in UB is there a definition of the phrase “value to society.” Let me propose a definition: the value of any clinical research study to society depends on how much that study can contribute to improving medical decision making for patients like those in the research study.

Clearly, major influences on the value of a study are how serious and/or how widespread is the disorder under study, the consequences of inadequate intervention, and the feasibility, costs, and risks of adequate intervention—all clinical issues. Here, and in UB, the focus is only on design decisions. An invalid design that misleads clinical decision making (type I error), a valid design with inadequate power that leaves the research question unanswered (type II error), or a valid and powerful design that answers a clinically irrelevant research question (sometimes called type III error) are all of little value to society.

The RCT is one type of clinical research study, specifically meant to inform choice between treatments (treatment T v. control/comparison C) for patients with a certain disorder. For an RCT to have value to society, the research question must be important to clinicians, and the design decisions (sampling, measurement, design, analysis) must lead to an accurate and precise indication of the clinical effect size of T v. C, some indication of how much better patient outcomes would be if T were used rather than C in the population sampled.

To be specific about such an effect size, if one patient were sampled from that population and given T and another given C, what is the probability that the one given T would have an outcome clinically preferable to the one given C (symbolically T > C) compared to the opposite (C > T): the success rate difference (SRD)?2,3 If the outcome measure were univariate and binary (success/fail), SRD equals \( p_1 - p_2 \), the effect size used in UB. If the outcome measure were univariate and normally distributed, SRD = \( 2\Phi(d/\sqrt{2}) - 1 \), where \( \Phi() \) is the standard cumulative normal distribution and \( d \) a form of Cohen’s \( d \).4 SRD can be estimated in an RCT for any outcome measures provided only that one could use them to compare the outcomes of 2 patients and decide which (if either) is clinically preferable, a necessary condition for an RCT outcome measure. Now, with estimation of SRD as the goal, let us now reconsider the UB arguments.

The discussion in UB focused on a limited subset of such RCTs, only those when no proven intervention existed. The issue of placebo (C = P) or nocebo/null treatment (C = N) control groups is more heatedly debated when effective interventions exist, for then, using either P or N might be viewed as ethically questionable, as essentially withholding effective treatment from patients for the duration of the RCT.
absence of effective intervention, the only nonplacebo control considered in UB is N, either alone or in combination with P.

Consider another alternative: treatment as usual (C = TAU). In this case, each sampled patient is randomized to T or to the treatment he or she would have been given had he or she not participated in the RCT (TAU). Patients in the TAU group would agree to have their response to treatment assessed according to the same protocol as used for T, preferably by “blinded” examiners to avoid bias.

Then if clinicians treating TAU patients knew there was no effective treatment and generally took a “wait-and-see” approach, TAU would be essentially equivalent to N. If clinicians mistakenly thought there were effective treatments and used them, TAU would be more effective than either C or P. If clinicians think what they are doing is effective, delivering the message that T is better than withholding such treatment (C = P or N) is unlikely to influence medical decision making.

There are, of course, objections to TAU as the control/comparison group, whether or not there are known effective treatments. However, the SRD comparing outcomes of T v. TAU would give an answer closest to what clinicians need to know: is T likely to be better than what we are doing now, and by how much?

In designing an RCT, the choice of C and the choice of appropriate outcome measure(s) are separate decisions. Whatever the outcome measures selected, they can be used with whatever C is selected. However, SRD will differ depending on which C is coupled with which outcome measures. Moreover, which combination is chosen has a major influence on the necessary sample size to ensure adequate power. Power does not determine the right research question. Choosing a design so as to minimize necessary sample size often leads to type III error. Instead, the right research question determines which design decisions are feasible and valid, and power then determines which of these would be most cost-effective to address that research question.5

As to choice of outcome measure, I agree with UB that the value of an RCT to society depends on assessing benefit-to-harm balance. However, the benefit to any individual patient may be multivariate (remission, reduction of symptoms, etc), and each variable may be binary, ordinal, or measured on a continuum. The same is true for harm (death, occasional rash or headache, sexual dysfunction, etc). Focusing on a single binary measure of benefit and a single binary measure of harm, as UB does, is limited, although useful for a demonstration.

So let us focus on the case used by UB. Consider the following three situations in Table 1, describing results obtained with 3 different interventions in the same population. In all 3, the probability of benefit (B+) is 50%; the probability of harm (H+) is 30%.

| Scenario a | B+ 0 | H– 50 | Row Total 50 |
| Scenario b | B+ 15 | H– 35 | Row Total 50 |
| Scenario c | B+ 30 | H– 20 | Row Total 50 |

Scenario a: benefit and harm negatively correlated; scenario b: benefit and harm uncorrelated; and scenario c: benefit and harm positively correlated.

In short, when the benefit-to-harm balance is considered as they affect individual patients, the correlation between benefit and harm (e.g., situations a v. b v. c) and the degree to which harm cancels benefit...
or benefit cancels harm when these co-occur within individual patients must be considered, for that has the potential to totally change recommendations to clinicians. UB dismiss these issues, pointing out (correctly) that considering outcome in this way is not current practice. However, every new solution to an existing methodological problem (including the UB proposal) is, by definition, not current practice, but if that solution is feasible and increases the value to society, it should be given serious consideration for future practice.

The crucial messages of UB are these: in considering the design of any RCT:

- The potential value to society should be considered. Potential value is not determined by \( P \) values but by effect sizes that reflect the potential impact on medical decision making.

Crucial in determining the value to society are the following:

- The choice of control/comparison group(s)
- Benefit-to-harm balance

I would only add the following:

- The messages of UB extend to any clinical research, far beyond RCTs in which no effective treatment is known and far beyond RCTs in general (to risk evaluations, to medical test or diagnosis development, etc).
- The most appropriate choice of outcome should reflect the benefit-to-harm balance for individual patients in the population sampled, not for the population as a whole or by considering benefits and harms separately. The societal value of medical decision making is determined by the potential effect on individual patients, who simultaneously experience both harms and benefits.

REFERENCES

1. Uyei J, Braithwaite RS. Are there scenarios when the use of non-placebo-control groups in experimental trial designs increase expected value to society? Med Decis Making. 2015;XX:[IN PRESS].