Double-Crossed: Why Crossover in Clinical Trials May Be Distorting Medical Science

Vinay Prasad, MD

Randomized controlled trials, particularly in cancer medicine, have increasingly allowed for “crossover.” Crossover refers to a study design in which patients not receiving the investigational agent are offered the treatment, typically after a set time, if their disease worsens. As clinical trial enrollment has diminished, crossover is thought to appeal to candidate patients and bolster recruitment.1

Crossover provides some theoretical advantages in randomized studies. Not only is there a control group, but patients serve as their own control. This advantage is more pertinent for trials evaluating subjective outcomes than for those studying objective ones (such as mortality). Moreover, for drugs that may impact survival, crossover permits ultimately beneficial drugs to be offered to more trial participants; even patients in the control group have a path to receive the novel agent.

At the same time, permitting crossover can also have serious disadvantages. The most concerning of these is that crossover may distort the outcomes of clinical studies. Recent trial history illustrates how crossover at time of progression may mask not only real survival benefits—as is commonly asserted—but also may mask real null or negative overall survival (OS) effects. In the case of the latter, crossover may mislead us regarding the real efficacy of an intervention.

For example, the inhibition of poly(adenosine diphosphate-ribose) polymerase (PARP), a regulator of the DNA base excision–repair pathway, has emerged as a promising therapeutic target in triple-negative breast and ovarian cancers. The PARP inhibitor olaparib was recently tested as maintenance therapy in a randomized phase II study involving patients with relapsed serous ovarian cancer who had at least a partial response to their most recent platinum therapy.2 The trial—which was not a crossover study—found that although the drug improved progression-free survival (PFS; 4.8 vs 8.4 months), this improvement did not translate into an OS benefit. For this reason, and because further analysis showed little likelihood that the drug would be able to improve overall survival, the manufacturer announced that it would not take the drug forward to phase III.3

Contrast the case of olaparib with that of another PARP inhibitor, iniparib, which was tested in a phase II randomized trial involving patients with metastatic triple-negative breast cancer. In the iniparib study, the drug was given along with chemotherapy (intervention group) and compared with chemotherapy alone. Crossover was permitted if disease progression occurred. The trial showed improvement in not only PFS (3.6 vs 5.9 months) but also OS (median, 7.7–12.3 months), both of which were highly significant.

Here we have the case of 2 PARP inhibitors. Both improve PFS by approximately 3 or 4 months. In one case, this improvement translated into OS benefit—nearly double the improvement in PFS—and in the other case, it didn’t. Further, the drug’s manufacturer then deemed it ineffective. Are these medications simply biologically different? Or is the difference in results due to a difference in trial design?

Allowing crossover may have led to the difference in OS in the iniparib study, not through the drug’s benefit, but by harming the control group.1 In the case of iniparib, the drug is thought not to exert antitumor effects by itself, but to potentiate the effects of the 2 chemotherapeutic agents. When resistance to chemotherapy occurs (progression), crossover to chemotherapy and a potentiator (iniparib) is unlikely to
afford much response or benefit. However, for patients who begin with the novel agent, progression removes them from the study and means that they can seek out alternative agents or trials that potentially may be more effective.

To further illustrate this point, consider the case of vandetanib, a controversial drug recently approved to treat late-stage medullary thyroid cancer. Approval was granted based on a single, non-crossover trial involving 331 patients, which allowed for open-label use on progression. That trial showed a benefit in PFS but absolutely no difference in OS. The drug’s toxicity—pneumonia, sepsis, respiratory failure, arrhythmia, and cardiac failure—is alleged to offset the tumor progression advantage, leading to no net benefit. Thus, what this medication’s effect on OS would be if open-label use were not allowed is altogether unclear. The benefit in PFS might remain, but deaths in the treatment arm may have led to worse overall mortality. Even if the trial merely found no improvement in OS, use of the drug could be questioned. Why accept real toxicity (and pay thousands of dollars) if gains made in preventing death from cancer were lost due to deaths from other causes?

We can thus evoke a general principle explaining where crossover will fail: any drug that beneficially affects a key surrogate measure (eg, progression) but has off-target effects that worsen overall outcomes will show false benefits if crossover is permitted. The harms will be borne by both groups, but the benefits detected in only one.

A final thought experiment furthers this argument. Imagine if the cholesterol esterase transfer inhibitor torcetrapib were tested in a crossover design. This is a medication that shows favorable effects on high-density lipoprotein for patients already on a statin but also shows a subtle increase in death from all causes (0.5% increased absolute risk of death after nearly 2 years). How would allowing crossover in the torcetrapib trial for patients who did not reach lipid targets affect the results? In the case of torcetrapib, detecting the harms, which were unknown and unpredictable, required a large number of patients (>7500) and almost 2 years’ follow-up. Crossover would have hidden the drug’s harm. Of course, lipid targets are not as well accepted as PFS as a surrogate, but PFS has also been criticized as not constituting a meaningful end point. And, as the examples illustrate, the general principle stands, regardless of surrogate used.

A final consideration is how crossover may distort subjective outcomes. Consider a recent study of duloxetine for chemotherapy-induced neuropathy. In this trial, patients with cancer who had painful neuropathy were randomized to a treatment arm receiving duloxetine for 6 weeks, a 1-week washout, and then placebo for 6 weeks, or another arm with the reverse order—a double crossover. The study reported positive findings at a national meeting. However, we can wonder if the benefit of the medication were simply active placebo. In that case, the drug—which had real side effects—led to favorable symptom scores simply because patients knew they were receiving active agent at that point. A single trial following these patients for 12 or 24 weeks might show regression to the mean, with the 2 groups indistinguishable by the end of the trial—the characteristic pattern of placebo response. Instead, because of double crossover, we are not afforded any long-term follow-up. The clinician’s question—should I prescribe this drug to my patient long term—because neuropathy rarely resolves—is not adequately answered, nor can it be. Crossover eliminates long-term follow-up.

Thus, crossover has the potential for and is possibly already distorting medical science. Empirical analysis of this issue is deeply limited, because trials are rarely performed on the same question, using or not using this technique. Although improving trial enrolment remains a noble endeavor, we cannot afford to sacrifice trial validity. The burden of proof must be on trialists who favor crossover to show scientific
Crossover in Clinical Trials

equivocality, a standard currently not met. At a minimum, the results of crossover trials, particularly concerning mortality, must be scrutinized for hidden harm signals.

References