Surrogate End Points
Are They Meaningful?

Two studies in this issue of the Journal, one by Um- menhofer et al.\(^1\) and the other by Furst and Rodarte,\(^2\) examine whether the antiemetic ondansetron decreases the incidence of postoperative nausea and vomiting in children. Both studies effectively demonstrate a desirable efficacy of this treatment: a decreased incidence of postoperative vomiting, a finding likely to bolster the use of ondansetron by anesthesiologists. Both studies define “success” as a reduced incidence of vomiting. I question, however, whether this is a reasonable end point.

As clinicians, our goal is to maintain patient safety, prevent adverse events, and improve outcome. All clinicians would likely agree that nausea and vomiting are undesirable events after surgery; some might categorize these as adverse outcomes. Yet simply counting the number of episodes of vomiting is not the only manner in which to evaluate efficacy and may not be the most appropriate. For example, Weinstein et al.\(^3\) recently demonstrated that children given morphine had a greater incidence of nausea and vomiting than those given a placebo and concluded that morphine induces vomiting. However, children given morphine may be more comfortable during recovery and thus more likely to ambulate and play. Movement, in turn, might trigger vomiting (a well-described phenomenon). Therefore, the greater incidence of vomiting in children given morphine may result from an indirect effect of morphine—improved analgesia—rather than a direct emetic effect. Similar arguments may apply to ondansetron. First, the decreased incidence of vomiting may result from an indirect effect. Second, does a decreased incidence of vomiting translate to improved clinical status or greater patient satisfaction?

More appropriate end points for antiemetic studies exist. Both Ummenhofer et al.\(^1\) and Furst and Rodarte\(^2\) argue that a potential benefit of ondansetron would be a decreased incidence of unplanned hospital admission after planned outpatient surgery, but neither group elected to study this end point. Similarly, patient satisfaction was not evaluated in these studies or in several other studies\(^4-14\) that demonstrate a reduced incidence of postoperative nausea and vomiting in response to ondansetron. In fact, most studies of this therapy (including that by Ummenhofer et al.\(^1\)) do not report the duration of recovery room stay,\(^4-6,14\) and none documents a shorter recovery period,\(^2,5\) even though this would be an excellent indicator of the cost-effectiveness of the treatment. For an adult in our institution, the least cost for 4 mg intravenous ondansetron is approximately $17; in turn, our pharmacy charges the patient $121. Will a patient or his or her insurance company willingly pay the increased charge for a medication that might decrease the number of episodes of nausea and vomiting (which, although undesirable, probably have no long-term adverse effects) but not decrease the duration of recovery room stay (and does not thereby offset the cost of the drug) and may or may not improve patient comfort (important, but difficult to quantify)? Under capitated care, will hospital administrators support any increased cost without clear-cut evidence that other costs can be reduced? Thus, studies using appropriate end points remain necessary to demonstrate the efficacy of ondansetron and of other new drugs.

End points other than those truly of interest (such as the incidence of vomiting rather than patient satisfaction or economic gain) are termed “surrogate” end points. Surrogate end points are used in many studies, sometimes because of historical precedent (as I believe is the case for most antiemetic studies), or because use of the true end point would require a massive sample. For example, a recent study comparing cardiac morbidity with various anesthetic techniques used as its measure of outcome the incidence or duration of myocardial ischemia.\(^15\) Undoubtedly ischemia is undesirable, but if it is transient and not associated with any further events, is it an appropriate outcome measure? The most appropriate outcome measure might be the incidence of myocardial infarction or congestive heart failure or the need for (or duration of) postoperative intensive care. Yet the incidence of these outcomes is so low that a massive (and prohibitively expensive) study would be necessary to demonstrate a desirable
effect. For example, if a new therapy decreased the incidence of postoperative myocardial infarction from 2% to 1%, more than 5,000 subjects would be required to achieve statistical significance (with a power of 80%). Because such massive studies are performed rarely (if ever), we often accept surrogate outcome measures rather than insist on true outcomes.

Surrogate outcomes appear throughout the anesthesia literature, and I too have used them. One good example is the demonstration that desflurane is associated with a more rapid time to initial responsiveness (e.g., eye-opening, response to simple commands, or answering of questions). Yet is it important whether a patient opens his or her eyes 10 min after completion of an anesthetic? This outcome is a surrogate for other, true outcomes such as the time at which a patient's trachea can be extubated (thereby permitting transfer from the operating room) or the time at which the patient can be discharged from the hospital.

I encourage clinical investigators to examine their outcome measures. Investigators heeding the suggestion to select appropriate outcome measures may be forced to alter their study designs (for example, by increasing sample size) to achieve more meaningful outcome measures. Studies with better outcome measures are more convincing and thus more important.

Dennis M. Fisher, M.D.
Department of Anesthesia
University of California, San Francisco
521 Parnassus Avenue
San Francisco, California 94143-0648

References


