The Role of Worst-Case Scenarios in Sponsors' and Regulators' Decision Making

In new drug development, 4 important considerations are acceptable safety, statistically significant efficacy, clinically significant efficacy, and manufacturability on a commercial scale. As it is possible for a drug candidate's efficacy to be statistically significant but not clinically significant, this editorial first focuses on the latter of these two considerations. It then discusses the issue of acceptable safety. The intent is to provide readers with a high-level overview of the role of worst-case scenarios in decision making as facilitated by the calculation of confidence intervals placed around a best-guess estimate from a single clinical trial.

Writing on the topics of statistical and clinical significance, Gardner and Altman¹ commented that “presenting p-values alone can lead to them being given more merit than they deserve. In particular, there is a tendency to equate statistical significance with medical importance or biological relevance,” something that must not be done. When deciding whether or not a drug candidate will receive marketing approval, consideration of statistical significance is necessary but not sufficient. Therefore, in cases where statistically significant evidence of efficacy is present, the next step is an evaluation of clinical significance.

When presenting the results of pivotal trials, both to regulators in marketing applications and to the medical community in journal publications, evaluation of clinical significance is typically performed via the placement of confidence intervals around the drug's treatment effect (mean effect in the drug treatment group minus mean effect of the control treatment), which is then referred to as the treatment effect point estimate. Confidence intervals facilitate quantification of the degree of confidence that is placed in the treatment effect point estimate.

When a randomized clinical trial is conducted, the group of individuals participating in the trial is a sample chosen from the population of all individuals with the disease or clinical condition of interest. While analysis of the trial's data provides a precise result for that particular sample, had a different sample been chosen the chances of the data obtained being identical to those obtained in the trial is so infinitesimally small that it can be meaningfully said that they would be different. The question of importance is this: How different might they likely be? Ideally, we would like them to be very similar since the greater the degree of similarity the greater the confidence we can reasonably place in the results from the trial.

While the word “confidence” in the previous sentence occurs in its everyday use, the term is also used in the discipline

of Statistics in a precise manner, just as the everyday word “significant” is used in a precise manner when describing a statistically or clinically significant result. A 2-sided confidence interval constitutes a range of values that are defined by the lower limit and the upper limit of the interval. In the current context, these limits lie equidistantly on either side of the treatment effect point estimate. The confidence interval is defined as a range of values that is likely to cover the true but unknown population treatment effect with a specified degree of certainty based on the data collected in the single clinical trial. That is, the employment of confidence intervals allows the extrapolation of observations from the single trial and its sample of participants to the population of interest.

Confidence Intervals

Various 2-sided confidence intervals (CIs) are employed in the literature, including 90%, 95%, and 99% CIs. Placing a confidence interval around a treatment effect point estimate allows the following statement to be made, where “XX%” can be replaced by 90%, 95%, or 99%:

- The 2-sided XX% CI placed around the treatment effect point estimate obtained from this single trial is a range of values that we are XX% confident will cover the true but unknown population treatment effect.

The widths of the 3 CIs placed around the same treatment effect point estimate will differ: the 90% CI will be the narrowest, the 95% CI will be wider than the 90% CI and narrower than the 99% CI, and the 99% CI will be the widest. To move from having 90% confidence that the range of values between the lower and upper limits of the interval covers the true but unknown population treatment effect to having 95% confidence that it does so, the range of values must be greater. The same logic applies when moving from 95% confidence to 99% confidence.

Employment of Confidence Intervals in Assessment of Clinically Significant Efficacy

Imagine that a clinical trial has been conducted for a new antihypertensive treatment in which the control treatment was a placebo. (Contemporary research into antihypertensive treatment often involves multidrug combination therapy and the control treatment may not be a placebo, but let's stick with our example solely for illustrative purposes.) The drug's treatment effect, calculated as mean blood pressure reduction in the drug

treatment group minus mean blood pressure reduction in the placebo treatment group, was 10 mmHg. This is now referred to as the treatment effect point estimate. The calculations performed to generate the chosen 2-sided 95% CI yield a value that is subtracted from the treatment effect point estimate to give the lower limit and added to the treatment effect point estimate to give the upper limit. Imagine that this value is 1.5 mmHg. The interval therefore has a lower limit of 8.5 mmHg and an upper limit of 11.5 mmHg. The result can be written as follows:

- The treatment effect point estimate and 2-sided 95% CI = 10.0 (8.5, 11.5)

The result allows the following statement to be made:

- The data obtained from this single clinical trial are compatible with a treatment effect in the general population of hypertensive individuals as small as 8.5 mmHg and as large as 11.5 mmHg, and our best estimate is 10.0 mmHg.

The question now is this: Does this result provide compelling evidence of likely clinically significant efficacy, that is, clinically significant therapeutic benefit? The process of determining clinical significance is not as formulaic as determining statistical significance: in addition to the statistical analysis just described it requires skilled clinical judgment. This judgment focuses on the lower limit of the confidence interval as this is the lowest estimate of benefit produced by our calculations. Using the nomenclature used in the Editorial's title, this is the worst-case scenario. Skilled clinical judgment is brought to bear, and a decision will be made whether or not a decrease in blood pressure of 8.5 mmHg is clinically significant. The clinicians involved in the drug's development program may decide that the answer is yes, and hence deem the drug's efficacy to be clinically significant in their judgment. If the drug is also considered acceptably safe, this may lead to a decision to submit an application for regulatory approval. If so, clinicians at the regulatory agency will also perform a skilled clinical judgment that will inform their marketing approval decision.

Now consider a different scenario. Imagine that the results from the same trial had been partially similar, but also notably different: the treatment effect point estimate was the same, but the lower and upper limits of the confidence interval were 2.0 mmHg and 18 mmHg, respectively. This result would be written as follows:

- The treatment effect point estimate and 2-sided 95% CI = 10.0 (2.0, 18.0)

The result allows the following statement to be made:

- The data obtained from this single clinical trial are compatible with a treatment effect in the general population of hypertensives as small as 2.0 mmHg and as large as 18.0 mmHg, and our best estimate is 10.0 mmHg.

Again, the question of importance is: Does this result provide compelling evidence of clinically significant efficacy? First, consider the treatment effect point estimate and the upper limit of the confidence interval. The treatment effect point estimate is our best estimate of the likely treatment effect in the general population, and it is exactly the same as in the previous case, i.e. 10 mmHg. The upper limit represents the greatest likely population treatment effect compatible with the results obtained from the clinical trial, and hence it represents the best-case scenario. This value is actually considerably higher than in the previous case, that is, 18.0 mmHg as compared with 11.5 mmHg. However, the key value is the worst-case scenario as represented by the lower limit. This value is considerably lower than in the previous case, that is, 2.0 mmHg as compared with 8.5 mmHg. Skilled clinical judgment will again be brought to bear, and a decision will be made as to whether or not a decrease in blood pressure of 2.0 mmHg is deemed clinically significant. While any degree of reduction in high blood pressure is theoretically desirable, a drug that may (only) reduce blood pressure by 2.0 mmHg in patients if it were to receive marketing approval may not be considered a useful drug by the sponsors' clinicians since there are already drugs approved that lower blood pressure by considerably greater amounts. If they came to that decision, they might decide to investigate using a higher dose of the drug, or to terminate the drug's development program and transfer the sponsor's resources to pursuing development of other drugs.

Employment of Confidence Intervals in Assessment of Acceptable Safety

I've written about the term *acceptable safety* in various forums.^{2,3} It may initially seem a strange one, but it is a legitimate one: it is an unfortunate but immutable fact that no pharmacologically active compound that provides therapeutic benefit can be guaranteed to have no risks. Benefit and risk considerations come together when assessing acceptable safety. If a drug dramatically reduces the chance of a heart attack but also increases the chance of mild headaches, this balance may be considered well worthwhile and the drug therefore deemed acceptably safe by many stakeholders. In contrast, if a drug reduces the chance of mild headaches but also dramatically increases the chance of a heart attack, the benefit-risk balance may well be considered unacceptable, that is, the drug is not deemed to be acceptably safe by many stakeholders. Of course, no increase in the risk of side effects is ideal. However, in the treatment of a serious disease, sponsors, regulators, and patients and their physicians may consider an increase of a certain magnitude in the risk of certain side effects to be acceptable.

One approach to the assessment of acceptable safety involves the use of relative risk ratios. Using the data from a pivotal trial, the number of adverse events of interest in the drug treatment group (the numerator in the ratio) can be compared with the number in the control treatment group (the denominator in the ratio). If there are exactly the same number

in each treatment group, the relative risk ratio will be unity, that is, 1.00. If there are more in the drug treatment group, the value will be greater than 1.00. (If there are actually less in the drug treatment group, the value will be less than 1.00, and there will be some degree of evidence that the drug is actually therapeutically beneficial in this regard.)

Consider a trial involving a drug for a disease of considerable concern for which there are currently few treatment options. While finding a drug with therapeutic benefit is therefore of great interest, we also need to consider risks of serious side effects such as heart attacks. Imagine that a relative risk ratio was computed for occurrences of heart attacks, and the value was 1.40. This value is now regarded as the relative risk ratio point estimate. A 2-sided 95% CI is computed, and the lower and upper limits are 1.10 and 1.75. (In the case of analyses involving ratios, the lower and upper limits do not lie equidistantly around the point estimate: the upper limit lies further away.) This result is written as follows:

- The relative risk ratio point estimate and 2-sided 95% CI = 1.40 (1.10, 1.75).

The result allows the following statement to be made:

- The data obtained from this single clinical trial are compatible with as little as a 10% increase and as much as a 75% increase in the risk of heart attacks in the general population, and our best estimate is an increase of 40%.

In contrast to our discussions of estimating clinical significance, where the confidence limit of interest was the lower limit, here the worst-case scenario is represented by the upper limit, that is, an increased risk of a heart attack in the general population of 75%. The next step involves interpretation of these results in conjunction with the results of efficacy analyses, and again skilled clinician judgment will play a role. Attention will also be paid to the best estimate available of increased absolute risk. Going from a 1 in 10 chance of the adverse event to a 1.75 in 10 chance (an increase of 75%) may be viewed very differently from going from a 1 in 1,000,000 chance of the adverse event to a 1.75 chance in 1,000,000 even though this too represents a 75% increase. Given the degree of therapeutic benefit found in the trial, it will be considered whether or not this increase in the risk of a heart attack is deemed acceptable.

Various stakeholders will be involved in determinations here. Regulators have to make decisions at the public health level, and may decide to approve the drug while also noting that patients with the disease of interest who are at higher risk of heart attacks even without taking this drug should engage in careful discussions with their physicians about the potential use of the drug. It is quite possible that some patients will decide that, for them, the drug is acceptably safe, while others will decide that it is not.

The Confidence Interval Limit of Primary Interest

From the perspective of regulatory decision making, aimed at protecting and promoting public health, it is the worst-case scenario that is always of primary interest.⁴ This is true for both safety and efficacy considerations. In contrast to this similarity, the difference is that the worst-case scenario is represented by different parameters in safety and efficacy considerations. In safety considerations, the worst-case scenario is the greatest likely degree of risk, captured by the upper limit of the relevant confidence interval. This limit represents the top end of the “risk scale,” that is, how great the risk might be in the population of patients who could receive the drug if approved. In considerations of clinically significant efficacy, the worst-case scenario is the smallest likely degree of therapeutic benefit in the population of patients who could receive the drug if approved, captured by the lower limit of the relevant confidence interval.

The Roles of Statistical Science, Clinical Science, and Regulatory Science in New Drug Development

New drug development requires attention be paid to ethical, intellectual, scientific, biological, statistical, clinical, regulatory, financial, legal, congressional, and social considerations, and this list is likely far from exhaustive.⁵ This Editorial has focused on statistical, clinical, and regulatory considerations with regard to worst-case scenarios and their role in decision making. It is hoped that readers will gain an appreciation of the importance of this triumvirate in assessments of clinically significant therapeutic benefit and acceptable safety.

J. Rick Turner, PhD, DSc
Editor Emeritus,

Therapeutic Innovation & Regulatory Science

References

1. Gardner MF, Altman DG. Estimation rather than hypothesis testing: confidence intervals rather than *p*-values. In: Gardner MF, Altman DG, eds. *Statistics With Confidence*. London: British Medical Association, 1986.
2. Turner JR, Karnad DR, Kothari S. *Cardiovascular Safety in Drug Development and Therapeutic Use: New Methodologies and Evolving Regulatory Landscapes*. Gewerbestrasse, Switzerland: Springer International Publishing, 2017.
3. Turner JR. Integrated cardiovascular safety: multifaceted considerations in drug development and therapeutic use. *Expert Opin Drug Saf*. 2017;16:481-492.
4. Turner JR. *Key Statistical Concepts in Clinical Trials for Pharma*. New York: Springer, 2012.
5. Turner JR. *New Drug Development: An Introduction to Clinical Trials*. 2nd ed. New York: Springer, 2010.