

### **EXHIBIT 3**

#### **STATISTICAL SIGNIFICANCE.**

The analysis and presentation of data collected from clinical trials (and other forms of biological, medical and public health research) is governed by the principles of biostatistics. A key question that arises in interpreting the results of a clinical trial is whether a given result – for example, a numerical difference in the incidence of a particular adverse event in two arms of a clinical trial – has statistical weight or is due to chance. The field of biostatistics provides the quantitative basis for understanding such data, and its principles ensure that the conclusions drawn from a study are valid and credible. Two statistical concepts that are used to evaluate the import of a given finding and that appear throughout this report – “statistical significance” and the “confidence interval” – are discussed below.

#### **A. Statistical Significance.**

In general terms, a given clinical trial result (for example, a difference in incidence rates) is considered “statistically significant” if it is unlikely to have occurred as a matter of chance – in other words, if it is unlikely that a difference in incidence rates between or among the arms of a trial would have been observed if the incidence rates in the corresponding populations as a whole were equal. The likelihood that a given result is due to chance is expressed in terms of a probability value, or “p-value.” The probability threshold below which a given result can no longer be attributed to chance is called the level of significance. A common choice for the level of significance is 5% or

0.05. Thus, generally, if the p-value for a given result – e.g., an observed difference in adverse event incidence rates between treatment group A and treatment group B – is less than or equal to 0.05, then the result is considered statistically significant and may be accorded weight.<sup>1</sup> If the p-value is greater than 0.05, the result is not statistically significant, and it may be assumed that no such differential outcome between treatment groups A and B would occur in the broader population from which the study subjects were drawn.<sup>2</sup>

Although all findings with p-values lower than 0.05 are considered statistically significant, p-values closer to zero are accorded greater weight than p-values closer to 0.05. In the presentation of clinical trial results generally and throughout this report, a p-value of 0.001 (for example) is expressed simply as  $p=0.001$ .

Although the p-value is a well-established measurement of statistical significance, it cannot be interpreted in isolation. If a given result is not statistically significant (i.e., p-value exceeded 0.05), one cannot dismiss it without first asking whether the study at

---

<sup>1</sup> P. Armitage\*, G. Berry\*, J.N.S. Matthews\*. *Statistical Methods in Medical Research*. 4<sup>th</sup> Edition, Blackwell Science, 2002, at 88.

<sup>2</sup> Robert W. Fletcher\* and Suzanne W. Fletcher\*. *Clinical Epidemiology*, 4<sup>th</sup> Edition, Lippincott Williams & Wilkins, 2005, at 172. More specifically, in designing a hypothesis-testing clinical trial, biostatisticians set up a so-called “null hypothesis” that assumes that the drug being tested has no effect on the disease-state at issue. For example, in a trial designed to prove the hypothesis that patients on Vioxx will experience fewer gastrointestinal adverse events than patients on naproxen, the null hypothesis would be that Vioxx and naproxen cause equal numbers of such adverse events. The statistical significance of results favoring Vioxx is then evaluated in relation to the null hypothesis. Study results suggesting that Vioxx causes fewer gastrointestinal adverse events than naproxen become statistically significant when the difference in the number of such adverse events on Vioxx and on naproxen becomes sufficiently extreme – i.e., sufficiently far outside the range of expected chance variations around an equal mean – that the null hypothesis must be rejected and the observed difference in effect assumed to be real. The point at which the null hypothesis must be rejected is set at  $p<0.05$  – that is, when the probability value of the differential in effect having been due to chance is less than 0.05.

issue had adequate “statistical power” to detect that result. The “statistical power” of a study is the probability that, assuming a given result is “real” – i.e., that the result reflects a biological reality in the population from which the study participants were drawn – the study at issue would have declared that result to be statistically significant (i.e., at a significance level of 0.05 or lower).

Power values exceeding 80 or 90% are usually considered sufficient. Power increases as a function of the sample size (number of subjects participating in a given study). Large clinical trials are generally designed and “powered” to detect effects constituting the primary and secondary endpoints of the trial but may not have sufficient power for post-hoc analyses or subgroup analyses. The smaller the number of patients at issue, the less likely it is that a treatment effect – even a biologically real effect – will rise to the level of statistical significance.<sup>3</sup> As a result, for tests with low power, such as, for example, the test of proportionality of hazard rates, biostatisticians may set the significance level above  $p=0.05$ .

B. Confidence Intervals.

Another way of expressing the weight to be accorded to a finding is by means of a “point estimate” and surrounding “confidence interval.” A point estimate is the statistically best “guess” or estimate of the true effect size being measured in a study – e.g., the true difference between incidence rates for patients on treatments A and B. The

---

<sup>3</sup> Conversely, the fact that a result is statistically significant does not necessarily mean that it has clinical significance. A p-value of 0.001, for example, conveys a high degree of confidence that an observed between-group difference is real but says nothing about the magnitude of that difference or whether it has any clinical impact. Robert W. Fletcher\* and Suzanne W. Fletcher\*. *Clinical Epidemiology*, 4<sup>th</sup> Edition, Lippincott Williams & Wilkins, 2005, at 172.

surrounding confidence interval is a range of values likely to encompass the actual effect size. The probability that a given confidence interval captures the true effect size is referred to as confidence level. A common choice for the confidence level is 95%, which is interpreted as follows:

If the study is unbiased, there is a 95% chance that the [confidence] interval includes the true effect size. The narrower the confidence interval, the more certain one can be about the size of the true effect. The true value is most likely to be close to the point estimate, less likely to be near the outer limits of the [confidence] interval, and could (5 times out of 100) fall outside these limits altogether.<sup>4</sup>

Confidence intervals, like p-values, reflect the statistical significance (or lack thereof) of a given finding. If the confidence interval includes a value corresponding to the null effect (no effect) – e.g., a zero difference in incidence rates of adverse events for patients on treatments A and B – the results are considered not statistically significant. If the value corresponding to the null effect – e.g., a risk difference of 0 or risk ratio of 1.0 – is not included in the range of the 95% confidence interval around the observed effect, the results can be interpreted as statistically significant at the 0.05 level.

---

<sup>4</sup> Robert W. Fletcher\* and Suzanne W. Fletcher\*. *Clinical Epidemiology*, 4<sup>th</sup> Edition, Lippincott Williams & Wilkins, 2005, at 172.

Like p-values, confidence intervals are influenced by sample size. In large studies, the confidence interval is narrow, which means the true effect can be pinpointed with greater accuracy. In a small study, the confidence interval is much wider, and the point estimate has less weight.<sup>5</sup>

\* \* \*

---

<sup>5</sup> David A. Savitz\*. *Interpreting Epidemiologic Evidence: Strategies for Design and Analysis*. Oxford University Press, 2003, at 257