Considerations for Using Positive Controls in Phase 2 Clinical Trials of Central Nervous System Disorders

Peer-reviewed author version

Mallinckrodt, Craig H.; Detke, Michael J.; Prucka, William R.; Ruberg, Stephen J. & MOLENBERGHS, Geert (2010) Considerations for Using Positive Controls in Phase 2 Clinical Trials of Central Nervous System Disorders. In: DRUG INFORMATION JOURNAL, 44(4). p. 431-441.

DOI: 10.1177/009286151004400407
Handle: http://hdl.handle.net/1942/11028

Considerations for Using Positive Controls

In Phase II Clinical Trials of Central Nervous System Disorders

Craig H. Mallinckrodt, Ph.D.[1], Michael J. Detke, M.D., Ph.D.[1,2],

William R Prucka, Ph.D.[1], Stephen J. Ruberg, Ph.D.[1],

and Geert Molenberghs, Ph.D.[3]

1. Eli Lilly & Co., Lilly Corporate Center, Indianapolis, IN 46285

2. Departments of Psychiatry, Harvard Medical School, Boston MA; McLean Hospital, Belmont MA; and, Indiana University School of Medicine, Indianapolis, IN

3. I-BioStat, Hasselt University, , Diepenbeek, Belgium, and, Katholieke Universiteit Leuven, Leuven, Belgium.

Abstract

Active comparators are often included in Phase II studies as a positive control to assess assay sensitivity, but inclusion of a positive control does not necessarily improve decision making. Simulation studies and illustrations are used to show that positive controls are more useful in assessing assay sensitivity as the probability the test drug is effective decreases and as power for the contrast of the positive control versus placebo increases. These results suggest that positive control should be powered at a minimum of 80%, and preferably at 90%. Analogously, a positive control can be used in an estimation framework to assess whether the study "performed as expected," thereby indicating whether or not the test drug was assessed under the anticipated conditions. In so doing, the sample would need to be sufficiently large to ensure reliable estimation. The key point again being that results of the positive control must be reliable if they are to be useful, and adding a small sample of patients in a positive control arm can do more harm than good. It is also important to recognize that including a positive control only allows assessment of assay sensitivity. Actual clinical-trial data are used to suggest that two smaller two-arm studies of test drug and placebo instead of one larger study that also includes a positive control may improve assay sensitivity with little to no increase in the total sample size.

Key words: Clinical trials, positive controls, power

Introduction

At present, only about 9% of central nervous system (CNS) drugs that enter phase I testing survive to launch [1]. Approximately 50% of the failures stem from failures to demonstrate efficacy in Phase II, which is a 15% increase in failure rate over the previous decade [1]. Meanwhile, failure rate of CNS drugs in Phase III is about 50% [2], with problems in drug-placebo discrimination and increased placebo response rising at an alarming rate [4]. Together, these findings clearly point to high rates of false negative and false positive rates in phase II as a major obstacle in CNS drug development. In fact, improving Proof-of-Concept (PoC) clinical trials is the most important factor required to improve the attrition rate in drug development [2].

Active comparators may be included in addition to a test drug and placebo in a clinical trial as a positive control to assess assay sensitivity; that is, to determine whether the study provided a valid test of the experimental drug [4]. Intuitively, inclusion of a positive control in addition to placebo should help foster better decisions from a PoC trial. However, most of the literature on this topic is in reference to including placebo in addition to an active drug so as to establish equivalence of the test and active drugs [4-12]. The merits of an active comparator as a positive control regarding inference of the test drug versus placebo have not been explored extensively in the literature regarding CNS drugs in general and psychiatric drugs in particular.

Hence, even though intuition suggests a positive control can foster better interpretations of the experimental drug, empirical data are lacking. Moreover, including a positive

control adds cost, time, and complexity to the study. Therefore, the objective of this research is to evaluate the operational characteristics of positive control arms that are necessary for inclusion of this arm to aid decision making in PoC clinical trials. This objective is approach through the use of numerical analysis of hypothetical examples, clinical trial simulation, and examination of actual clinical-trial data.

Framework and Context

Describing the results of a study in only a hypothesis testing framework (Was the difference statistically significant or not?) ignores potentially important distinctions. For example, consider three studies where the criterion for success was based on a statistically significant difference between test drug and placebo on the primary analysis based on a p-value cut off of .05. If the p-values for these three studies were .049, .051, and .999, based solely on statistically significant study 1 looks very different from studies 2 and 3, whereas studies 2 and 3 look similar. In reality, studies 1 and 2 are very similar, it just so happens that the p-values fall on either side of the arbitrary cut off of .05; and in reality, although neither studies 2 or 3 were significant, in study 2 the difference barely missed significance whereas in study 3 drug was almost certainly not superior to placebo.

Despite the loss of information inherent to dichotomizing p-values as significant or not, considering results in this black and white framework is a useful starting point for understanding the role of positive controls when assessing assay sensitivity.

As will become clear in subsequent sections, the utility of using a positive control depends in part on whether or not the test drug is effective. Therefore, the use of positive controls to assess sensitivity is first addressed separately for cases where the test drug is in truth effective and when it is not effective. This distinction is useful to first fix ideas, but also results in hypothetical assessments, because efficacy of the test drug is typically not known at the start of Phase II. This difficulty is resolved in a subsequent section by moving from the black and white framework of effective or not effective to a probability of success framework. In addition, given the limitations of the black and white world of the hypothesis testing framework, a subsequent section also considers positive controls in an estimation framework.

Scenarios Where the Test Drug is Effective

The following tables summarize the probabilities of various outcomes when a test drug and a positive control are simultaneously compared with placebo in a study. Results are described in terms of success and failure, based on presence or absence of a statistically significant difference from placebo. Understanding these probabilities is the first step in developing a framework for quantitatively evaluating the utility of a positive control in phase II studies.

Table 1 depicts a scenario where the test drug is effective, with the test drug and the positive control (correctly) powered at 80%. In this scenario, the correct result would be for both drugs to yield a significant difference from placebo. However, simply due to chance alone each drug is expected to be non-significant in 20% of the trials. Hence,

assuming independence of outcomes, an assumption explored momentarily, 64% (.80 * .80) of the trials are expected to yield a significant difference for both the test drug and the positive control. Therefore, at least one "wrong" result is expected in 36% (100% - 64%) of the trials. More specifically, in 16% (.8 * .2) of the trials only the test drug is expected to be significant. In another 16%, (.2 * .8) only the positive control is expected to be significant, and in 4% (.2 * .2) neither is expected to be significant.

Separation from placebo of the test drug by definition demonstrates assay sensitivity. Hence, in the 80% of the trials where the correct result of test drug separating from placebo is obtained, decision making is neither hindered nor helped by having the wrong or correct result for the positive control. However, in the 20% of the trials where the wrong result is obtained for the test drug, in the 16% of trials where the positive control separates, the presence of the positive control reinforces the false negative result because it is logical to believe that the test drug should have separated if it were effective, because the study was sensitive to the effects of the positive control. In such cases, development would probably be stopped and the negative result would never be discovered as a false negative. In the 4% of trials where neither arm is significant, the positive control helps inform us that the result for the test drug is a false negative, potentially leading to another study to further evaluate the drug.

Table 1. Probabilities of outcomes when the test drug is effective and the positive control and the test drug are powered at 80% - assuming independence of outcomes.

	Significance of Positive control	Total

		Yes	No	
Significance of Test drug	Yes	64	16	80
	No	16	4	20
Total		80	20	

Table 2 is the same as Table 1 except that now each drug is powered at 50% rather than 80%, mimicking scenarios such as depression where effective drugs have been shown to separate from placebo only about half the time regardless of sample sizes [13]; similar problems in drug-placebo discrimination may be evolving in schizophrenia [14]. Alternatively, such a scenario may arise in disease states where assay sensitivity is not a problem, but fewer patients are randomized to the positive control than to the test drug.

The correct result would again be for both drugs to yield a significant difference from placebo. However, as a consequence of lower power, only 25% of the trials are expected to yield a significant difference for both drugs. Therefore, at least one "wrong" result is expected in 75% of the trials. More specifically, in 25% of the trials only the test drug is expected to be significant, in 25% only the positive control is expected to be significant, and in 25% both are expected to be insignificant.

As before, in the trials where the correct result is obtained for the test drug, decision making is neither hindered nor helped by having the wrong or correct result for the positive control. However, this occurs only 25% of the time. In the 50% of trials where the wrong result is obtained for the test drug, the 25% of trials where the positive control separates the presence of the positive control reinforces the false negative result; and in

the 25% where neither is significant, the positive control helps confirm that the result for the test drug is a false negative.

Table 2. Probabilities of outcomes when the test drug is effective and the positive control and the test drug are powered at 50% - assuming independence of outcomes.

		Significance of Positive control		Total
		Yes	No	
Significance of Test drug	Yes	25	25	50
	No	25	25	50
Total		50	50	

For simplicity, Table 2 and Table 3 were developed assuming independence of outcomes for the test drug and positive control. However, the assumption of independence is not justified because each treatment is compared with the same placebo arm. Since the current scenario has concordant true effects of the two treatments, a higher degree of concordance in results is expected than would be obtained from independent outcomes.

Table 3 depicts the results from a simulation study where each treatment arm had approximately 82% power. The observed frequencies showed a greater percentage of concordant results than the expected frequencies assuming independence (in parentheses). Specifically, concordant results were obtained in approximately 80% (71.4% + 7.8%) of the outcomes whereas, as with independence, 70% (66.9 + 3.3) concordant outcomes were expected. The basic pattern seen in Table 3 is similar to Table 1. However, the increase in concordance from non-independence reduced the frequency of trials where

inclusion of a positive control reinforced a false negative result, but increased what may be viewed as a confusing finding in which neither the experimental nor positive control separated from placebo.

Table 3. Probabilities of outcomes when the test drug is effective and the positive control and the test drug are powered at 80% - not assuming independence - outcomes determined by simulation.

		Significance of Positive control		Total
		Yes	No	
Significance of Test drug	Yes	71.4 (66.9)	10.6 (15.1)	82.0
	No	10.2 (14.7)	7.8 (3.3)	18.0
Total		81.6	18.4	

Validate, move to production, and rerun pgm positive_control_1. Update results

Scenarios Where the Test Drug Is Not Effective

Table 4 depicts a scenario where the test drug is not effective (test drug equal to placebo), with the test drug tested at alpha = .10, and the positive control (correctly) powered at 80%. In this scenario, the correct result would be for the positive control to yield a significant difference and for the test drug to be insignificant. However, simply owing to chance alone, the test drug is expected to be significant in 10% of the trials (with alpha = .10 the false positive rate is 10%) and the positive control is expected to be insignificant in 20% of the trials. Hence, assuming independence of outcomes, an assumption explored momentarily, 8% (.1 * .8) of the trials are expected to yield a significant difference for both drugs, and the test drug only is expected to be significant in 2% of trials. The correct result (test drug not significant and positive control

significant) is expected in 72% of trials. In 18% of the trials neither drug is expected to separate from placebo.

Therefore, at least one "wrong" result is obtained in 28% of the trials. When the wrong result is obtained for the test drug (10% or the trials), presence of the positive control does not help identify the result as a false positive and a large scale Phase III development program would likely follow, only to see that program fail. In the 18% of the trials where neither drug separates the presence of the positive control suggests the trial was failed and thereby failed to adequately evaluate the test drug, erroneously suggesting another proof of concept study is needed. In the 72% of trials where the correct result is obtained, the positive control reinforces the belief that the test drug is ineffective.

Table 4. Probabilities of outcomes when the test drug is not effective, with alpha set and .10 and the positive control powered at 80% - assuming independence of outcomes.

		Significance of Positive control		Total
		Yes	No	
Significance of Test drug	Yes	8	2	10
	No	72	18	90
Total		80	20	

Table 5 depicts a scenario similar to Table 4 except the positive control is powered at 50% rather than 80%, as might be the case when randomizing fewer patients to the positive control than to the test drug, or in disease states where known active compounds

frequently fail, even with large sample sizes. Again, the correct result would be for the positive control to yield a significant difference and for the test drug to be non-significant. However, simply owing to chance alone, the test drug is again expected to be significant in 10% of the trials, but the positive control is expected to be insignificant in 50% of the trials. Hence, assuming independence of outcomes, 5% of the trials are expected to yield a significant difference for both drugs, and the test drug only is expected to be significant in 5% of trials. The correct result of the test drug being non-significant with positive control significant is expected in 45% of the trials. In 45% of the trials, neither drug is expected to be significant.

Therefore, at least one "wrong" result is expected in 55% of the trials. When the wrong result is obtained for the test drug (10% of the trials), presence of the positive control does not help identify the result as a false positive and a large scale Phase III development program would likely follow, only to see that program fail. In the 45% of the trials where neither drug separates, the presence of the positive control suggests the trial was failed and thereby failed to adequately evaluate the test drug, erroneously suggesting another study is needed. In the 45% of trials where the correct result was obtained, the presence of the positive control correctly reinforces the belief that the test drug is ineffective.

Table 5. Probabilities of outcomes when the test drug is not effective, with alpha set and .10 and the positive control powered at 50% - assuming independence of outcomes.

	Significance of Positive control	Total

		Yes	No	
Significance of Test drug	Yes	5	5	10
	No	45	45	90
Total		50	50	

For simplicity, Table 4 and Table 5 were developed assuming independence of outcomes for the test drug and positive control arms. However, the assumption of independence is not justified because all treatment arms are compared with the same placebo arm. Hence, some deviation in results from assuming independence is expected; however, the true effects of the test drug and positive control are not concordant, and it is not intuitively obvious how results should vary from those expected assuming independence.

Table 6 depicts the results from a simulation study where the test drug did not differ from placebo and the positive control had approximately 80% power. Comparisons of the observed frequencies with the expected frequencies assuming independence (in parentheses) show similar results. Therefore, results from the test drug and positive control were reasonably independent; and, results in either Table 4 or Table 6 reasonably reflect what to expect when a test drug is not effective and a positive control is powered at 80%.

Table 6. Probabilities of outcomes when the test drug is not effective, using alpha - .10, and the positive control is powered at 80% - not assuming independence - outcomes determined by simulation.

	Significance of Positive control	Total

		Yes	No	
Significance of Test drug	Yes	7.3 (8.1)	3.0 (2.2)	10.3
	No	71.7 (70.8)	18.0 (18.9)	89.7
Total		79.0	19.0	

Validate, move to production, and rerun pgm positive_control_1. Update results

---

Probability-of-success Framework

The summaries in the previous sections. assuming a drug is or is not effective are intuitively useful, but of limited use in that if it were known whether or not the test drug was effective, there would be no need to do the study. Therefore, it is useful to consider positive controls from a perspective based on the probabilities of the test drug being effective.

To incorporate this perspective, the consequences of the various outcomes on drug development decisions and the utility of the positive control in making those decisions needs to be considered. Assume that, if the test drug is significant, development will proceed to Phase III; if the test drug is not significant and the positive control is significant, then development is stopped; if neither the test drug nor the positive control is significant, then the PoC trial is repeated.

To classify positive control outcomes as neutral, helpful or harmful, consider first each quadrant of Table 3 and then Table 6, with focus on these tables because they do not assume independence and therefore yield slightly more accurate probabilities. When the

test drug is in fact effective, positive controls have a neutral effect whenever the test drug separates from placebo (left and right quadrant of the top row of Table 3) because the correct answer was obtained for the test drug and the positive control does not add information. When the effective test drug fails to separate, the positive control is harmful when it separates from placebo (lower left quadrant of Table 3) because it reinforces the false negative result by suggesting the study was capable of finding a difference in the test drug if a difference existed. The positive control is helpful when it does not separate from placebo (lower right quadrant of Table 3) because it suggests that the study did not adequately evaluate the test drug.

The results from Table 3 (test drug effective) are repeated in Table 7 with additional information about the decision, utility of the positive control, and action taken. The probabilities that the positive control has a helpful, neutral, or harmful effect are 0.078, 0.820, and 0.102, respectively. Thus, in these scenarios where the test drug is effective, the positive control is seldom beneficial and is just as often harmful.

Table 7. Probabilities of outcomes and their consequences when the test drug is effective and the positive control and the test drug are powered at 80% - not assuming independence.

Active Outcome	Test Outcome	Result / Decision	Probability	Action	Utility of Active arm	Cost / gain
Y	Y	True positive	0.714	Proceed to Phase III	Neutral	Unneeded arm in PoC study
Y	N	False negative	0.102	Kill drug	Harmful	Opportunity for new treatment lost
N	Y	True positive	0.106	Proceed to Phase III	Neutral	Unneeded arm in PoC

						study
N	N	No decision	0.078	Repeat PoC	Helpful	Opportunity for new treatment preserved

The results from Table 6 (test drug not effective) are repeated in Table 8 with additional information about the decision, utility of the positive control, and action taken. When the test drug separates from placebo, the positive control reinforces the false positive result (left and right quadrants of the top row in Table 6). When the test drug does not separate from placebo and the positive control does (left quadrant of the lower row in Table 6) ,the positive control helps because it provides more confidence that the study was capable of finding an effect of the test drug if an effect existed. When neither drug separates from placebo (lower right quadrant of Table 6), the positive control is harmful because it erroneously suggests that the study was not capable of finding a difference. The probabilities that the positive control has a helpful or harmful effect are 0.717 and 0.283, respectively, with the positive control never having a neutral effect. .

Table 8. Probabilities of outcomes and their consequences when the test drug is not effective and the positive control is powered at 80% - not assuming independence

Active Outcome	Test Outcome	Result / Decision	Probability	Action	Utility of Active arm	Cost
Y	Y	False Positive	0.073	Proceed to Phase III	Harmful	Expensi ve, futile Phase III
Y	N	True Negative	0.717	Kill drug	Helpful	None
N	Y	False	0.030	Proceed to	Harmful	Expensi

		Positive		Phase III		ve, futile Phase III
N	N	No decision	0.180	Repeat PoC	Harmful	Repeat futile PoC

It is also necessary to factor in the probability of technical success p(TS) rather than separately considering scenarios for effective and ineffective test drugs. To incorporate p(TS), two scenarios are considered: p(TS) = 50% and 35%.

Table 9 summarizes the utility of positive controls across a portfolio of compounds when the positive control is powered at 80%. These results are based on the probabilities as shown in Tables 7 and 8. Table 10 provides the same information based on powering of the positive control at 90%.

When the positive control is powered at 80% , with p(TS) = 50%, the frequency of the positive control having a beneficial, neutral, or harmful effect across the portfolio is 40%, 41%, and 19%. When powered at 80% with a lower p(TS) of 33% the corresponding percentages are 51%, 27%, and 22%.

When the positive control is powered at 90%, with p(TS) = 50%, the frequencies of a positive control having a beneficial, neutral, or harmful effect are 42%, 45%, and 13%, respectively. When powered at 90% with a lower p(TS) of 33% the corresponding percentages are 55%, 30%, and 15%.

Table 9. Frequencies of neutral, helpful, and harmful outcomes of positive controls when powered at 80%[1].

	Percentage of Trials	
	p(TS) = 50%	p(TS) = 33.3%
Positive control Neutral	41%	27%
Positive control Helped	40%	51%
Positive control Hurt	19%	22%

1. Assume an infinitely large number of compounds are tested, with the percentage of compounds that are effective defined by p(TS), and the percentage of ineffective compounds defined by 1 – p(TS).

Table 10. Frequencies of neutral, helpful, and harmful outcomes of positive controls when powered at 90%[1].

	Percentage of Trials	
	p(TS) = 50%	p(TS) = 33.3%
Positive control Neutral	45%	30%
Positive control Helped	42%	55%
Positive control Hurt	13%	15%

1. Assume an infinitely large number of compounds are tested, with the percentage of compounds that are effective defined by p(TS), and the percentage of ineffective compounds defined by 1 – p(TS).

Estimation Framework

Rather than basing assay sensitivity on a hypothesis test to ascertain whether the known effective positive control separated from placebo, an estimation framework can be used. For example, instead of looking at statistical significance, the point estimate and associated confidence intervals for the contrast of the positive control versus placebo can be compared with historical data to see whether the positive control results are in line with historical results.

Although both the hypothesis testing and estimation frameworks rely on the same data, the estimation framework provides a more granular look at the results, and therefore has potential advantages. Details on estimation versus hypothesis testing are beyond the scope of this text. Instead, we focus on what questions need to be addressed to understand the utility of a positive control to assess assay sensitivity because both frameworks lead to the same fundamental point. Namely, if results from the positive control are to be beneficial in assisting decision making, they must be reliable enough to be trusted.

In the hypothesis testing framework reliability can be addressed via power, whereas in the estimation framework reliability can be addressed via width of the confidence interval for the positive control contrast with placebo.

A Real-data Example

Prior to approval of duloxetine for major depressive disorder, 11 clinical trials were conducted that included 15 treatment arms of duloxetine tested versus placebo. In 7 of those studies, which included 11 duloxetine treatment arms, a positive control (SSRI) was also included. Among these 7 SSRI arms, 5 had equal randomization to duloxetine and 2 had half as many patients as duloxetine. These studies have been published individually [15-22] and in summaries [23,24,] with additional details being available at Lillytrials.com [25] .

If each duloxetine arm is viewed as a stand alone PoC trial, the value of the positive control in regards to assessing assay sensitivity can be evaluated. Results from the 11 duloxetine arms are cross tabulated with results from the SSRI arms in Table 11.

Duloxetine separated from placebo in 8/11 (73%) contrasts. SSRI separated from placebo in 2/7 (28%) contrasts. In the 8 contrasts where duloxetine separated from placebo (top left and top right quadrant of Table 11), including the SSRI had a neutral effect. In the 3 contrasts where duloxetine did not separate, the SSRI helped in 1 instance because neither duloxetine nor SSRI separated from placebo (lower right quadrant of Table 11), suggesting a failed study; however, in 2 contrasts SSRI erroneously supported the false negative result for duloxetine (lower left quadrant of Table 11).

Table 11. Results from duloxetine and SSRIs from clinical trials in major depressive disorder.

		Significance of SSRI Positive control	
		Yes	No
Significance of Duloxetine	Yes	4	4
	No	2	1

Effect sizes from the primary analysis of the HAMD17 total score and whether or not that difference was statistically significant, are summarized by dose in Table 12. Interestingly, duloxetine separated from placebo in 4/4 (100%) of the studies that did not include a positive control. The average effect size on the HAMD17 from two arm studies was 0.53 compared with 0.39 for studies that included a positive control. Previous research has

shown that subscales of the HAMD may improve signal detection compared with the total score [26-29]. The average effect size on the Maier subscale [30] of the HAMD from two arm studies was 0.63 compared with 0.45 for studies that included a positive control.

The larger effect sizes from the two-arm studies are especially noteworthy in that the average duloxetine daily dose in those studies was just over 60mg, whereas in the studies with a positive control the average daily dose was approximately 80 mg. Moreover, these results are consistent with a recent report showing that the percentage of patients randomized to placebo was the most influential factor in determining drug-placebo discrimination in antidepressant clinical trials [31].

It is also noteworthy that 12 of the 15 duloxetine treatment arms arose from having 2 identical studies run via the same protocol. Each study was independently and adequately powered, but designed to be pooled to increase precision. At least 1 positive result was obtained at each dose level in each pair of studies.

Table 12. Effect sizes and statistical significance from the primary analysis of duloxetine clinical trials in major depressive disorder[1].

Trials with a positive control

Trial	40	60	80	120
HMAT-A	.249		.272	
HMAT-B	**.378**		**.567**	
HMAY-A			**.490**	**.726**
HMAY-B			**.302**	**.359**
HMAQ-A				**.520**
HMAQ-B				.150

HMCR		**.273**		
Average	.314	.273	.401	.439

Trials that did not have a positive control

HMBH-A	**.720**
HMBH-B	**.320**
HMNB	**.520**
HQAC	**.550**
Average	**.528**

[1] Effect sizes from contrasts that were significantly different are in bold type.

Validate, move to production, and rerun pgms dulox1 and dulox2. Cross reference appropriate WPDF for effect size spreadsheet

Discussion

While it may be tempting to assume including a positive control is useful given that it allows assessment of assay sensitivity, the example scenarios, data simulations, and real-data examples included in this research suggest the matter needs to be evaluated on a case-by-case basis and is not clear cut.

Careful consideration reveals the following conundrum. If assay sensitivity is low, it is difficult to trust the results of the study, especially negative results for a test drug. But including a positive control may not improve decision making since the results from this contrast with placebo are also unreliable. On the other hand, if assay sensitivity is good,

results of the study can be trusted and we can trust the results from the test drug, thereby negating the need for a positive control.

More specifically, in scenarios where an effective test drug separates from placebo, the inclusion of a positive control is not useful because separation of the test drug establishes assay sensitivity. When the effective test drug fails to separate, the frequency of the positive control reinforcing a false negative result versus correctly identifying a failed study is proportional to the power of the active arm. In other words, if the test drug is effective, the probability of good decision making stemming from use of a positive control is ***inversely*** related to the power of the positive control.

In scenarios where the test drug is not effective, when the test drug is significant, inclusion of a positive control does not protect against type 1 error, but rather reinforces it. When the ineffective test drug fails to separate, the frequency that the positive control reinforces a true negative result versus incorrectly identifying a failed study is proportional to the power of the active arm. In other words, if the test drug is not effective, the probability of a correct decision stemming from use of a positive control is ***directly*** related to power of the positive control.

Taking a probability perspective on the use of positive controls to assess assay sensitivity reveals that the positive control should be highly powered, perhaps 80%-90%, for it to be useful. In other words, results of the positive control must be reliable if they are to be useful. In addition, the positive control will be more useful than when p(TS) is lower.

It is important to realize that a positive control assesses assay sensitivity, but it does not improve assay sensitivity. In fact, a recent meta-analysis demonstrated that one of the most important factors influencing drug-placebo discrimination was the percentage of patients randomized to placebo [31]. As the percentage of patients randomized to placebo decreased, for example when adding a positive control, the drug-placebo difference decreased. Similar findings have been reported by other researchers 32-34 and are consistent with the duloxetine example presented in this paper.

Examination of the summary basis approval data set of antidepressant clinical trials showed that an active comparator separated from placebo in only about 60% of the trials [32], well below the 80%-90% required to enhance decision making. Therefore, rather than *assessing* assay sensitivity, means to *improve* assay sensitivity should be considered. For example, a second, independent study may provide greater protection against false negative results than including a positive control. Other means of improving signal detection of antidepressants, such as use of subscales rather than total scores on assessment scales, and better analytic methods, may be useful more broadly across psychiatric clinical trials.

References

1. Hurko, O and JL Ryan. Translational research in central nervous system drug discovery. Neurotherapeutics. 2005;2:671-682.

2. Kola, Ismail and John Landis. Can the pharmaceutical industry reduce attrition rates? *Nature Reviews Drug Discovery* **3**, 711-716 (August 2004) | doi:10.1038/nrd1470

3. Kenp, Aaron S. Nina R. Schooler, Amir H. Kalali, Larry Alphs, Ravi Anand, George Awad, Michael Davidson, Sanjay Dube´, Larry Ereshefsky, Georges Gharabawi, Andrew C. Leon, Jean- Pierre Lepine, Steven G. Potkin, and An Vermeulen. What Is Causing the Reduced Drug-Placebo Difference in Recent Schizophrenia Clinical Trials and What Can be Done About It? Schizophrenia Bulletin doi:10.1093/schbul/sbn110

4. Temple R, Ellenberg SS: Placebo-controlled trials and active-control trials in the evaluation of new treatments. 1. Ethical and scientific issues. Ann Intern Med 2000; 133:455–463

5. Tsong Y, Zhang J (2005). Testing superiority and non-inferiority hypotheses in active controlled active trials. Biometrical Journal, 47, 62-74.

6. Hung HMJ, Wang S-J, Tsong Y, Lawrence J, O'Neil RT (2003). Some fundamental issues with non-inferiority testing in active controlled trials. Statistics in Med., 22, 213-225.

7. ICH Efficacy Document No. E-9 (1997). Statistical Principles of Clinical Trials. http://www.fda.gov/cder/guidance.

8. DerSimonian R, Laird N (1986). Meta-analysis in clinical trials. CCT, 177-188.

9. Sankoh, AJ, Huque, MF. Impact of Multiple Endpoints on Type I Error Rate and Power of Test Statistic in Non-superiority Clinical Trials. Far East Journal of Theoretical Statistics, 13 (1), 47-65.

10. Sankoh AJ, Al-Osh M, Huque FM (1999). On the utility of the Dirichlet distribution for meta-analysis of clinical studies. JBS, 9: 289-306.

11. Rohmel J (1998). Therapeutic equivalence investigations: statistical consideration. Stats in Med, 17, 1703-1714

12. Holmgreen EB (1999). Establishing equivalence by showing that a specified percentage of the effect of the active control over placebo is maintained JBS 9(4), 651-659

13. Khan, A., Kolts, R.L., Thase, M.E., Krishnan, K.R., Brown, W., 2004b.Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. Am. J. Psychiatry 161, 2045–2049.

14. Kenp, Aaron S. Nina R. Schooler, Amir H. Kalali, Larry Alphs, Ravi Anand, George Awad, Michael Davidson, Sanjay Dube´, Larry Ereshefsky, Georges Gharabawi, Andrew C. Leon, Jean- Pierre Lepine, Steven G. Potkin, and An Vermeulen. What Is Causing the Reduced Drug-Placebo Difference in Recent Schizophrenia Clinical Trials and What Can be Done About It? Schizophrenia Bulletin
doi:10.1093/schbul/sbn110

15. Goldstein DJ, Mallinckrodt C, Lu Y, Demitrack MA: Duloxetine in the treatment of major depressive disorder: a double-blind clinical trial. J Clin Psychiatry 2002;63: 225-231.

16.Goldstein DJ, Lu Y, Detke MJ, Wiltse C, Mallinckrodt C, Demitrack MA: Duloxetine in the treatment of depression: a double-blind placebo-controlled comparison with paroxetine. J Clin Psychopharmacol 2004;24: 389-399.

17. Detke MJ, Wiltse CG, Mallinckrodt CH, McNamara RK, Demitrack MA, Bitter I: Duloxetine in the acute and long-term treatment of major depressive disorder: a placebo and paroxetine-controlled trial. Eur Neuropsychopharmacol 2004;14: 457-470.

18. Perahia DG, Wang F, Mallinckrodt CH, Walker DJ, Detke MJ: Duloxetine in the treatment of major depressive disorder: a placebo- and paroxetine-controlled trial. Eur Psychiatry 2006;21: 367-378.

19. Nierenberg AA, Greist JH, Mallinckrodt CH, Prakash A, Sambunaris A, Tollefson GD, Wohlreich MM. Duloxetine versus escitalopram and placebo in the treatment of patients with major depressive disorder: onset of antidepressant action, a non-inferiority study. Current Medical Research and Opinion. 2007;23(2):401-416.

20. Detke MJ, Lu Y, Goldstein DJ, Hayes JR, Demitrack MA: Duloxetine, 60 mg once daily, for major depressive disorder: a randomized double-blind placebo-controlled trial. J Clin Psychiatry 2002;63: 308-315.

21. Detke MJ, Lu Y, Goldstein DJ, McNamara RK, Demitrack MA: Duloxetine 60 mg once daily dosing versus placebo in the acute treatment of major depression. J Psychiatr Res 2002;36: 383-390.

22. Raskin J, Wiltse CG, Siegal A, Sheikh J, Xu J, Dinkel JJ, Rotz BT, Mohs RC: Efficacy of duloxetine on cognition, depression, and pain in elderly patients with major depressive disorder: an 8-week, double-blind, placebo-controlled trial. Am J Psychiatry 2007;164:900-909.

23. Nemeroff CB, Schatzberg AF, Goldstein DJ, Detke MJ, Mallinckrodt C, Lu Y, Tran

PV: Duloxetine for the treatment of major depressive disorder. Psychopharmacol Bull 2002;36: 106-132.

24. Mallinckrodt CH, Prakash A, Houston JP, Swindle R, Detke MJ, Fava M. Differential antidepressant symptom efficacy: Placebo-controlled comparisons of duloxetine and SSRIs (fluoxetine, paroxetine, escitalopram). Neuropsychobiology. 2007;56:73-85.

25. INSERT PROPER CITE FOR LILLY TRIALS

26. Bech, Per, Daniel K. KAJDASZ, and Vibeke Porsdal. Dose-response relationship of duloxetine in placebo-controlled clinical trials in patients with major depressive disorder**.** Psychopharmacologia. 2006, vol. 188, n°3, pp. 273-280 [8 page(s) (article)] (1 p.1/4)

27. Entsuah R, Shaffer M, Zhang J. A critical examination of the sensitivity of unidimensional subscales derived from the Hamilton Depression Rating Scale to antidepressant drug effects. Journal of Psychiatric Research 2002; 437-448.

28. Faries D, Herrera J, Rayamajhi J, DeBrota D, Demitrack M, Potter WZ. The responsiveness of the Hamilton Depression Rating Scale. Journal of Psychiatric Research 2000; 34:3-10.

29. Mallinckrodt, Craig H, Adam L. Meyers, Apurva Prakash, Douglas E. Faries, Michael J. Detke. Simple Options for Improving Signal Detection in Antidepressant Clinical Trials. 2007. PsychoPharm Bull. 40(2) 101-114.

30. Maier W, Philipp M. Improving the assessment of severity of depressive states: a reduction of the Hamilton Depression Scale. Pharmacopsychiatry 1985; 18:114-115.

31. Papakostas. George I. and Maurizio Fava. Does the probability of receiving placebo influence clinical trial outcome? A meta-regression of double-blind, randomized clinical trials in MDD. European Neuropsychopharmacology (2009) 19, 34–40

32. Khan A, Detke M, Khan SR, Mallinckrodt C. Placebo response and antidepressant clinical trial outcome. Journal of Nervous and Mental Disease 2003; 191:211-218.

33. Khan, A., Khan, S.R., Walens, G., Kolts, R., Giller, E.L., 2003. Frequency of positive studies among fixed and flexible dose antidepressant clinical trials: an analysis of the food and drug administration summary basis of approval reports. Neuropsychopharmacology 28, 552–557.

34. Khan, A., Kolts, R.L., Thase, M.E., Krishnan, K.R., Brown, W., 2004b.Research design features and patient characteristics associated with the outcome of antidepressant clinical trials. Am. J. Psychiatry 161, 2045–2049.