Document Title: **Draft Guidance for Industry on Non-Inferiority Clinical Trials** [Docket No. FDA–2010–D–0075]		**Date: 20-May-2010**

Page/ Line No	Comment type *	Comment (with rationale)	Proposed change
General comments			
	N	The document covered different aspects of non- inferiority clinical trials in particular for pharmaceutical products. Some other areas are not represented.	Some examples using biologicals (especially in vaccine trials) and/or safety could be included to enrich the text and showing broader applications.
	M	When refer to p-values, there is not always mention if this is a one sided or a two sided.	Please clarify
	N	The guidance is specially focused to establish a non inferiority margin in case of efficacy studies.	Some guidance on safety would be useful. In that case, M2 only is usually defined and can be moving according to the rate. More frequent event could lead to a larger clinical margin (referring to Kem Phillips method) Please consider to expand in some extent).
	N	To finalize the parallelism between superiority & non inferiority, the concept of M2 could be introduced in the superiority as the minimum clinical efficacy required. This concept is a standard in prophylactic vaccines	To add this concept in section III.A.1.
	N	The discussion of III.A.1 is overly complicated and not at all transparent to an un-initiated reader. Superiority is simply a special case of non-inferiority where M1=M2=0 and the NI alternative/null hypotheses are reversed. Both are valid approaches, but there has been a whole lot more superiority testing than NI testing historically.	To clarify this concept in section III.A.1.
	N	Guidance should be provided in this document on acceptable approaches to the design of NI studies with multiple primary endpoints. Use of the intersection-union procedure is ill-advised in many such settings.	Some guidance on this would be useful.
	N	Repeated references throughout the draft guidance are made to `assay' sensitivity when it is `trial' sensitivity that is meant (e.g. line 231). Assay sensitivity is the term used also in ICH-E10 and consistency of terminology is useful, however, it is easily confounded with the sensitivity of biological assays.	Please explained what is meant by assay sensitivity in this context.
	N	Little mention is given to the scale of the margin (e.g., multiplicative vs. additive) even though it also has large implications.	Some more guidance on this would be useful.

	N	Throughout the guideline it is assumed that the CIs of different parameters are always equally tailed. In the chapter “Logic of the NI Trial” the lower bound of the two-sided 95% CI is considered to be equivalent to the lower bound of the one-sided 97.5% CI. There are quite a lot of CI methods for which this is not true.	Please clarify.
	N	The document advocates that a placebo control should be included throughout. The sample size is also based on the difference between a current standard treatment and placebo. In most established diseases however, it is unethical to use a placebo, and current standard treatments may also not have been compared to placebo. The document comments that a placebo control is not always possible or ethical, however it does not advise on how to determine a sample size in the absence of a placebo.	
	M	The document is highly repetitive which makes it overly complicated and difficult to extract the relevant information. It would be more easily understood if it was condensed, possibly including a summary. Diagrams would be useful to help to understand examples better when discussing choosing the NI margin	
	N	There isn’t much guidance for the analysis of a NI trial other than to extend to test for superiority. Can this be added? There is a brief section on ITT and as-treated, but does it need to discuss the 1-sided strategy? Similarly the sample size requirements should change depending on whether you expect the new treatment to be the same, better or worse. This should be mentioned.	
	I	What randomisation ratio is recommended, 1:1 or uneven to allow more experience of the new product at the risk of increasing the sample size?	
	I	Why use a 95% CI? Tradition?	
	M	Why use a 2-sided CI throughout the guidance for a 1-sided problem? ICH E9 clearly states "For non-inferiority trials a one-sided interval should be used."	
	I	What considerations for the choice of power are recommended? It's surprising to see 80% used even when this implies a 1/5 chance of failing to show non-inferiority even when it may be highly likely it exists.	

	M	Who should approve M2 - HAs, ECs?	
	M	Why is bio-creep not mentioned? This is the gradual erosion of treatment effect by a series of trials where each time the new treatment is a bit worse than one in the previous study.	
	N	What about guidance for reporting non-inferiority studies? Often it is poor, with the pre-defined clinically relevant difference being omitted, not allowing the study's results to be interpreted, and only an irrelevant non-significant p-value being provided to "compare" treatment groups, with the wrong conclusion of "equivalence" from the under-powered study	
	M	The document should explain more clearly in which situations a test against M1 might be sufficient and in which situations a test against M2 must be performed in addition. The CHMP guideline on the choice of the non-inferiority margin clearly differentiates between trials aimed at indirectly showing superiority to placebo and trials aiming to show there is no important loss in efficacy if the new product is used instead of the control. The authors' focus appears to be on the first objective, but the second is mentioned as well and it is not clear enough which methodology best supports which objective.	Please clarify.
	I	The document should comment on the possibility to use continuous endpoints in an NI study.	Please clarify.
Specific comments			
Page 1 Footnote 2	I	The footnote refers to "therapeutic" biologic but does not include prophylactics	Eliminate the word "therapeutic" and replace by "biological products"
l.103	N	"In a placebo-controlled trial, the"	"In a placebo-controlled superiority trial, the"
Page 4 Figure 1	I	Figure 1 is not complete. There is always the possibility that T-P is either marginally inferior or statistically inferior to 0.	Please revise.
Page 4 Figure 1	I	Cases 2 and 3 are not really distinct as presented	Please clarify that in one case the estimated value is zero versus non zero, even if the lower limit of the CI is below zero in both situations.
Page 4 Line 123	I	The statement "study perhaps too small" is related with sample size and power and is not adding any value here	Delete "(study perhaps too small)"

Figure 2	N	The figure could be completed to cover the most frequent situations.	Add a case “7” in the figure where the point estimate lies between 0 and 1, the lower limit between -1 and 0 and the upper limit between 1 and 2
Figure 2	I	Figure 2 is not complete. There is always the possibility that C-T is indistinguishable or marginally above M1 but greater than 0 or that C-T statistically exceeds M1.	Please revise.
Page 5 Line 158	N	This outcome is not at all unusual, and the FDA guidance should provide guidance on how to interpret such results and how bio-creep to placebo will be avoided. Combination drug products would also be expected to show such characteristics if absorption were altered slightly by mixing of the separate drugs relative to separate administration, etc.	Consider to add this in the text
Page 5	N	It may be good to revisit fig 2 with M2 to ensure there is no misunderstanding. If M2 is specified there should be no interpretive problem for case number 6.	Consider to add this in the text
Lines 164-165	I	The sentence “It must be estimated (really assumed)” can be confusing	Replace by “it must be assumed based on …”
l.166	N	"Determining the NI margin is the single greatest challenge in the design, conduct, and interpretation of NI trials."	"Determining the NI margin is the single greatest challenge in the design of NI trials."
l.169	I	"The smaller the margin, the smaller the upper bound of the 95% two-sided confidence interval for C-T must be"	This doesn't consider other factors such as the power and, for binomial responses, the expected proportion or odds ratio which also influence the sample size.
l.179	I	"provides an important benefit"	the examples "life-saving or preventing irreversible injury" are ok but in other studies, the endpoint may be very different e.g. the patient may become pregnant. Moreover, such studies do not include placebo arms.
l.311	I	"As M2 represents a clinical judgment, there may be a greater flexibility in interpreting a 95% upper bound for C-T that is slightly greater than M2, as long as the upper bound is still well less than M1 (see Figure 3)."	This sounds like "if you just fail to meet your target, you may be ok after all"… so the predefined clinical judgment can be ignored. Examples 2 and 3 in Figure 3 are handled differently but who makes the decision?
Page 9 Line 326	I	It is not clear what the authors means by “but unacceptable loss of the control effect”.	Please clarify

Page 10 Line 341	I	'Note that AS is related to M1, our best estimate of the effect of the control in the study...' The best estimate would be the point estimate? Is this what is meant? Or does this mean the lower CI for M1? Or a value for M1 that was chosen even lower than this e.g. because the constancy assumption may not be valid?	M1 is used for slightly different things throughout the document, please clarify.
Page 10 Line 344	I	'..not have shown that M2 was ruled out' - should this read 50% loss was ruled out?	Please clarify.
Page 11 Line 378	I	Constancy assumption: is there any experience that can be shared about a possible impact of moving from a placebo-controlled to an active controlled study design? Could this be a reason to expect M1 to be bigger in the NI trial than in the placebo-controlled trial, e.g. for indications with subjective assessments?	Please clarify.
Page 12 Line 432	N	'bias towards the null'	Clarify that the null of a superiority trial is meant here (also applies to line 442)
l.439	M	"It should also be appreciated that intent-to-treat approaches, which preserve the principle that all patients are analyzed according to the treatment to which they have been randomized even if they do not receive it, although conservative in superiority trials, are not conservative in an NI study, and can contribute to this bias toward the null."	So should PP take precedence over ITT analyses in NI trials?
Page 12 Line 447	I	The concept of compliance be lead to misunderstandings	Please be more explicit; the poor compliance has a biased impact on non-inferiority. Does that mean that the margins should be reconsidered when the quality of the study is poor?
l.464	I	"The point estimate of the drug effect and its confidence interval (usually 95% but could be 90% or 99% under some circumstances) ..."	under what circumstances?
l.470	I	"Such similarity might be concluded, however, if the point estimate of the test drug favored it over the control and the upper bound of the 95% CI for C-T was close to showing superiority."	why should similarity be concluded if the test drug was favored and showed superiority? Surely it should be shown to be similar to the control, not better than it.
Page 16 Line 619	T	Choice of NI Method	Choice of NI Margin
l.634	I	"There was a clinical decision to ensure that not more than 50% of the effect of streptokinase was lost"	Here 50% is used as an example and earlier it was 40%, but in many cases, the maximum loss may be 20% or even less.

Page 18 Lines 723 and 727	N	95% CI	Please clarify whether this is one-sided or two-sided. Also applies to lines 768 and 1175.
Page 21 Line 856	N	'.. a representation of where the result is likely to be 95% of the time .. in a future study'.	Is this really true? This is an unusual interpretation of a confidence interval and may need some more explanation.
Page 22	N	Ref :"determining HESDE from single study vs. multiple trials". Multiples trials evaluation allows the opportunity to evaluate an overall estimate of the treatment effect of the active control as well as a measure of the study-to-study variability of that treatment effect.	This is also feasible in a single study by defining subgroup. This can be explained.
Page 22 Line 894	T	`can' should be replaced with `should'	Please correct.
Page 22 Line 905	N	References should be added for appropriate meta-analytic strategies.	Please add references.
Page 23 Point 3 Line 930	N	Pooling of large outcome studies; when saying it would be inappropriate to have the point estimate for one of these fall below the 95%CI lower bound of the pooled study data. CI for the considered point estimate of a single study could be much larger than CI for the pooled estimate and so this would be not justifiable to consider that point estimate as inappropriate	A clarification is needed.
Page 23 Point 4 Line 944	N	The authors suggest to use the largest effect (point estimate) regardless of the CI around that estimate. This is not in line with the 95%-95% or 90%-95% methods described on page 28	A clarification is needed.

Page 24 Section c	N	The authors suggest different metric for treatment effect. More sophisticated metric, for instance accounting for an effect that would depend linearly on the event rate in the control group, could be used.	Add a less traditional metric which led to the NI approach developed by Kem F. Phillips (Statist. Med. 2003; 22:201–212) could be mentioned
Page 26 Lines 1046 and 1048	T	Success results Success rates	Failure rates Failure rates
Page 28 Line 1163	I	“the study quality would not affect M2 when it is very small compared to M1”. Some precision should be given	What does “quality” means here? Compliance? What does “very small” indicates? Any idea of relative magnitude? Please clarify.
l.1258	N	"In the synthesis approach, the NI margin is not predetermined, but the outcome of the NI study, a consideration of the effect of the test agent vs. placebo, can be judged for adequacy."	Is this non-specification of the NI margin to be recommended? See also l.1287.
l.1287	N	"Clinical judgment is used to pre-specify an acceptable fraction of the control therapy’s effect that should be retained by the test drug, regardless of the magnitude of the control effect."	This clarifies l.1258 that something must be pre-specified.
l.1292	I	"so the clinical judgment to determine the choice of M2 is difficult and is generally not made until results are seen."	Does this mean the approach can't work?
l.1341	I	"A further problem is posed by the possibility that event rates will be lower in the new study."	This is not true when the expected rates are 50% but the observed rates are 45% - the actual power will be higher than hoped for.
l.1348	I	"A similar approach could be applied in an NI trial with upward adjustment of the sample size if the event rate is unexpectedly low."	Again, this example would be reversed in some situations.
l.1363	T	"Intent-to-treat)ITT) analyses"	"Intent-to-treat (ITT) analyses"
l.1372	M	"It is therefore important to conduct both ITT and as-treated analyses in NI studies."	This only partially addresses the problem. What about PP analyses?
Section G	N	Role of Adaptive Design in NI studies	To consider to move to the guidance for adaptive design.
Page 34 Line 1414	T	level of no more than 5% should be 2.5% (?) in addition to that line 1411 specify a two-sided CI ?	Please verify the levels across the document.

l.1443	I	"If the active control has shown superiority to other active treatments in the past, ..."	but what about if "If the active control has shown non inferiority to other active treatments in the past"? This leads to the question of bio-creep which surprisingly isn't mentioned at all in this guidance.
l.1481	I	"A typical value for M2 is often 50% of M1, at least partly because the sample sizes needed to rule out a smaller loss become impractically large."	This seems a strange argument for being flexible on M2
l.1483	N	"In this case, there is a better argument for some degree of flexibility if the study did not quite rule out the M2 margin; there might be reason to consider, for example, assurance of 48% retention (but not the expected 50%) for M2 as acceptable."	As stated above, it seems odd to be generous here but not for superiority.
Page 36 Lines 1513 to 1515	N	In which situation heterogeneity of covariate may increase the variance? Is this related with the use of random effects model?	This could be re-wording: situation where there is heterogeneity of group difference (or treatment effect) among covariate, the variance may be increased
l.1554	T	"it is important to determine that the previous conclusion"	"it is important to determine whether the previous conclusion"
l.1570	N	"Such a change might be sought because it would permit a smaller study or was more feasible given current event rates."	Again this seems to be a strange argument, possibly contrary to clinical judgment.
l.1765	T	"in this case 50% or one-half (versus placebo),"	"in this case 50% (versus placebo),"
l.2037	I	"*the event rate of either group needs further verification from each article"	is this part of the table or a comment in the guidance?
l.2085	I	"Study 2 rules out M1 using a fixed margin approach, but Study 1 does not."	is this true for Study 2 when the target was 1.09 and the upper limit was also 1.09?
Page 55 Line 2102	T	'no more than 90%'	'no more than 10%'?