
Guidance for Industry

Non-Inferiority Clinical

Trials

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact Robert Temple at 301-796-2270 or Robert O'Neill at 301-796-1700 (CDER), or the Office of Communication, Outreach, and Development (CBER) at 301-800-835-4709 or 301-827-1800.

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)**

**March 2010
Clinical/Medical**

Guidance for Industry

Non-Inferiority Clinical

Trials

Additional copies are available from:

Office of Communication

Division of Drug Information, WO51, Room 2201

Center for Drug Evaluation and Research

Food and Drug Administration

10903 New Hampshire Ave.

Silver Spring, MD 20993

Phone: 301-796-3400; Fax: 301-847-8714

druginfo@fda.hhs.gov

<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>

or

Office of Communication, Outreach, and Development

Center for Biologics Evaluation and Research

Food and Drug Administration

1401 Rockville Pike, Rockville, MD 20852-1448

Phone: 800-835-4709 or 301-827-1800

ocd@fda.hhs.gov

<http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/default.htm>

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
March 2010
Clinical/Medical

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
II.	BACKGROUND	1
III.	GENERAL CONSIDERATION OF NON-INFERIORITY STUDIES: REGULATORY, STUDY DESIGN, SCIENTIFIC, AND STATISTICAL ISSUES.....	2
A.	Basic Principles of a Non-Inferiority Study	2
B.	Practical Considerations in Use of NI Designs.....	13
IV.	CHOOSING THE NON-INFERIORITY MARGIN AND ANALYZING THE RESULTS OF AN NI TRIAL	17
A.	Introduction	17
B.	Statistical Uncertainties in the NI Study and Quantification of Treatment Effect of Active Control.....	19
C.	Statistical Methods for NI Analysis	27
D.	Considerations for Selecting M_2, the Clinical Margin, and the Role of Subjective Judgment.....	31
E.	Estimating the Sample Size for an NI Study.....	32
F.	Potential Biases in an NI Study	33
G.	Role of Adaptive Designs in NI Studies — Sample Size Re-estimation to Increase the Size of an NI Trial	33
H.	Testing NI and Superiority in an NI Study	34
V.	COMMONLY ASKED QUESTIONS AND GENERAL GUIDANCE	35
	APPENDIX — EXAMPLES.....	40
	REFERENCES - EXAMPLES	56
	GENERAL REFERENCES.....	59

Contains Nonbinding Recommendations

Draft – Not for Implementation

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37

Guidance for Industry¹ Non-Inferiority Clinical Trials

This draft guidance, when finalized, will represent the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. You can use an alternative approach if the approach satisfies the requirements of the applicable statutes and regulations. If you want to discuss an alternative approach, contact the FDA staff responsible for implementing this guidance. If you cannot identify the appropriate FDA staff, call the appropriate number listed on the title page of this guidance.

I. INTRODUCTION

This guidance provides sponsors and review staff in the Center for Drug Evaluation and Research (CDER) and Center for Biologic Evaluation and Research (CBER) at the Food and Drug Administration (FDA) with our interpretation of the underlying principles involved in the use of non-inferiority (NI) study designs to provide evidence of the effectiveness of a drug or biologic.² The guidance gives advice on when NI studies can be interpretable, on how to choose the NI margin, and how to analyze the results.

II. BACKGROUND

This guidance consists of four parts. The first part is a general discussion of regulatory, study design, scientific, and statistical issues associated with the use of non-inferiority studies when these are used to establish the effectiveness of a new drug. The second part focuses on some of these issues in more detail, notably the quantitative analytical and statistical approaches used to determine the non-inferiority margin for use in NI studies, as well as the advantages and disadvantages of available methods. The third part addresses commonly asked questions about NI studies and provides practical advice about various approaches. The fourth part includes five examples of successful and unsuccessful efforts to define non-inferiority margins and conduct NI studies.³

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidance describes the Agency's current thinking on a subject and should be viewed as recommendations unless specific regulatory or statutory requirements

¹ This guidance has been prepared by the Office of Biostatistics and the Office of New Drugs in the Center for Drug Evaluation and Research (CDER) and the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration.

² For the purposes of this guidance, all references to *drugs* include both human drugs and therapeutic biologic products unless otherwise specified.

³ References: in this guidance, reference to methods or studies are not included in the text; rather they are included in a General Reference section and a separate reference section for the examples in the Appendix.

38 are cited. The use of the word *should* in Agency guidances means that something is
39 suggested or recommended, not that it is required.

40

41

42 **III. GENERAL CONSIDERATION OF NON-INFERIORITY STUDIES:**
43 **REGULATORY, STUDY DESIGN, SCIENTIFIC, AND STATISTICAL**
44 **ISSUES**

45

46 **A. Basic Principles of a Non-Inferiority Study**

47

48 *1. Superiority Trials versus Non-Inferiority Trials to Demonstrate Effectiveness*

49

50 FDA’s regulations on adequate and well-controlled studies (21 CFR 314.126) describe four
51 kinds of concurrently controlled trials that provide evidence of effectiveness. Three of them
52 — placebo, no treatment, and dose-response controlled trials — are superiority trials that
53 seek to show that a test drug is superior to the control (placebo, no treatment, or a lower dose
54 of the test drug). The fourth kind of concurrent control, comparison with an active treatment
55 (active control), can also be a superiority trial, if the intent is to show that the new drug is
56 more effective than the control. More commonly, however, the goal of such studies is to
57 show that the difference between the new and active control treatment is small, small enough
58 to allow the known effectiveness of the active control to support the conclusion that the new
59 test drug is also effective. How to design and interpret such studies so that they can support
60 such a conclusion is a formidable challenge.

61

62 These active control trials, which are not intended to show superiority of the test drug, but to
63 show that the new treatment is not inferior to an unacceptable extent, were once called
64 equivalence trials, but this is a misnomer, as true equivalence (i.e., assurance that the test
65 drug is not **any** less effective than the control), could only be shown by demonstrating
66 superiority. Because the intent of the trial is one-sided (i.e., to show that the new drug is not
67 materially worse than the control), they are now called non-inferiority (NI) trials. But that
68 too, is a misnomer, as guaranteeing that the test drug is not any (even a little) less effective
69 than the control can only be demonstrated by showing that the test drug is superior. What
70 non-inferiority trials seek to show is that any difference between the two treatments is small
71 enough to allow a conclusion that the new drug has at least some effect or, in many cases, an
72 effect that is not too much smaller than the active control.

73

74 The critical difference between superiority and NI trials is that a properly designed and
75 conducted superiority trial, if successful in showing a difference, is entirely interpretable
76 without further assumptions (other than lack of bias or poor study conduct); that is, the result
77 speaks for itself and requires no further extra-study information. In contrast, the NI study is
78 dependent on knowing something that is not measured in the study, namely, that the active
79 control had its expected effect in the NI study. This is critical to knowing that the trial had
80 *assay sensitivity* (i.e., could have distinguished an effective from an ineffective drug). A
81 successful superiority trial has, by definition, assay sensitivity. A “successful” NI trial, one
82 that shows what appears to be an acceptably small difference between treatments, may or

163 The critical problem, and the major focus of this guidance, is determining M_1 , which is not
164 measured in the NI study (there is no concurrent placebo group). It must be estimated (really
165 assumed) based on the past performance of the active control and by comparison of prior test
166 conditions to the current test environment (see section III.A.4). Determining the NI margin
167 is the single greatest challenge in the design, conduct, and interpretation of NI trials.

168
169 The choice of the margin M_1 has important practical consequences. The smaller the margin,
170 the smaller the upper bound of the 95% two-sided confidence interval for C-T must be, and
171 the larger the sample size that will be needed.

172

173 3. Reasons for Using a Non-Inferiority Design

174

175 The usual reason for using a non-inferiority active control study design instead of a study
176 design having more readily interpretable results (i.e., a superiority trial) is an ethical one.
177 Specifically, this design is chosen when it would not be ethical to use a placebo, or a no-
178 treatment control, or a very low dose of an active drug, because there is an effective
179 treatment that provides an important benefit (e.g., life-saving or preventing irreversible
180 injury) available to patients for the condition to be studied in the trial. Whether a placebo
181 control can be used depends on the nature of the benefits provided by available therapy. The
182 International Conference on Harmonization guidance E10 on *Choice of Control Group and*
183 *Related Issues in Clinical Trials* (ICH E10) states:

184

185 In cases where an available treatment is known to prevent serious harm, such as death
186 or irreversible morbidity in the study population, it is generally inappropriate to use a
187 placebo control. [The term “generally” leaves room for a placebo control if the
188 known effective treatment is very toxic.]

189

190 In other situations, where there is no serious harm, it is generally considered ethical
191 to ask patients to participate in a placebo-controlled trial, even if they may experience
192 discomfort as a result, provided the setting is non-coercive and patients are fully
193 informed about available therapies and the consequences of delaying treatment.

194

195 There are, however, other reasons for using an active control: (1) interest in comparative
196 effectiveness and (2) assessing the adequacy (assay sensitivity) of a placebo-controlled study.
197 These are not the focus of this guidance, but will be considered briefly.

198

199 a. Comparative effectiveness

200

201 There is growing interest among third party payers and some regulatory authorities, on both
202 cost effectiveness and medical grounds, in the comparative effectiveness of treatments, and
203 an increasing number of such studies are being conducted. A critical issue is the importance
204 of including a placebo group, as well as the active comparator, in such studies (a 3-arm trial)
205 to assess assay sensitivity (i.e., the ability of the trial to detect differences of a specified size
206 between treatments). When the treatment is clinically critical, it will, of course, not be
207 ethically acceptable to include a placebo group, and the discussion of NI studies that follows
208 will be highly relevant to such trials. Even where it would be ethical to include a placebo

Contains Nonbinding Recommendations

Draft – Not for Implementation

209 group in addition to the active treatments (e.g., in studies of a symptomatic treatment), one is
210 not necessarily included in these comparative trials. Such omission of a placebo group may
211 render such studies uninformative, however, when they show no difference between
212 treatments, unless assay sensitivity can be supported in some other way.

213
214 Where comparative effectiveness is the principal interest, it is usually important—where it is
215 ethical, as would be the case in most symptomatic conditions—to include a placebo control
216 as well as the active control. Trials of most symptomatic treatments have a significant failure
217 rate (i.e., they often cannot show the drug is superior to placebo). Where that is the case in a
218 comparative trial, seeing no difference between treatments is uninformative. Inclusion of a
219 placebo group can provide clear evidence that the study did have assay sensitivity (the ability
220 to distinguish effective from ineffective treatments), critical if a finding of no difference
221 between treatments is to be interpretable. For example, we have seen that approximately
222 50% of all placebo-controlled antidepressant trials of effective agents cannot distinguish drug
223 from placebo. A trial in which two antidepressants are compared and found to have a similar
224 effect is informative only if we know that the two drugs can be distinguished from the
225 concurrent placebo group.

226

227 b. Assessing assay sensitivity of a placebo-controlled study

228

229 Although a successful superiority trial (e.g., placebo-controlled) is readily interpreted, a
230 failed trial of this design is not. Failure to show superiority to placebo can mean that the
231 drug is ineffective or that the trial lacked assay sensitivity. To distinguish between these two
232 possibilities, it is often useful to include an active control in placebo-controlled studies of
233 drugs in a class or condition where known effective drugs often cannot be distinguished from
234 placebo (e.g., depression, allergic rhinitis, angina, and many other symptomatic conditions).
235 If the active control is superior to placebo but the test drug is not, one can conclude that the
236 test drug lacks effectiveness (or at least is less effective than the active control). If neither
237 the active control nor the test drug is superior to placebo, the trial lacked assay sensitivity and
238 is uninformative about the effect of the test drug.

239

240 4. *The Non-Inferiority Margin*

241

242 As described above, the NI study seeks to show that the difference in response between the
243 active control (C) and the test drug (T), (C-T), the amount by which the control is superior to
244 test drug, is less than some pre-specified non-inferiority margin (M). M can be no larger than
245 the presumed entire effect of the active control in the NI study, and the margin based on that
246 whole active control effect is generally referred to as M_1 . It is critical to reiterate that M_1 is
247 not measured in the NI trial, but must be assumed based on past performance of the active
248 control, the comparison of the current NI study with prior studies, and assessment of the
249 quality of the NI study (see below). The validity of any conclusion from the NI study
250 depends on the choice of M_1 . If, for example, the NI margin is chosen as 10 (because we are
251 sure the control had an effect of at least that size), and the study does indeed rule out a
252 difference of 10 (seeming to demonstrate “effectiveness” of T), but the true effect of C in this
253 study was actually less than 10, say 5, T would not in fact have been shown to have any

Contains Nonbinding Recommendations

Draft – Not for Implementation

254 effect at all; it will only appear to have had such an effect. The choice of M_1 , and assurance
255 that this effect was present in the trial (i.e., the presence of assay sensitivity) is thus critical to
256 obtaining a meaningful, correct answer in an NI study.

257

258 Because the consequence of choosing a margin greater than the actual treatment effect of the
259 active control in the study is the false conclusion that a new drug is effective (a very bad
260 public health outcome), there is a powerful tendency to be conservative in the choice of
261 margin and in the statistical analysis that seeks to rule out a degree of inferiority of the test
262 drug to the active control of more than that margin. This is generally done by ensuring that
263 the upper bound of the 95% two-sided confidence interval for C-T is smaller than M_1 . The
264 upper bound of the confidence interval for C-T is not, however, the only measurement of
265 interest, just as the lower bound of a 95% confidence interval for effect size of drug versus
266 placebo is not the only value of relevance in a placebo-controlled trial. The point estimate of
267 the treatment effect and the distribution of estimates of C-T smaller than the 95% upper
268 bound are also relevant. Nonetheless, the upper bound of the 95% CI is typically used to
269 judge the effectiveness of the test drug in the NI study, just as a two-sided p-value of 0.05 or
270 less is traditionally the standard used for defining success in a superiority trial. The 95% CI
271 upper bound for C-T is used to provide a reasonably high level of assurance that the test drug
272 does, in fact, have an effect greater than zero (i.e., that it has not lost all of the effect of the
273 active control).

274

275 Although the NI margin used in a trial can be no larger than the entire assumed effect of the
276 active control in the NI study (M_1), it is usual and generally desirable to choose a smaller
277 value, called M_2 , for the NI margin. Showing non-inferiority to M_1 would provide assurance
278 that the test drug had an effect greater than zero. However, in many cases that would not be
279 sufficient assurance that the test drug had a clinically meaningful effect. After all, the reason
280 for using the NI design is the perceived value of the active control drug. It would not usually
281 be acceptable to lose most of that active control's effect in a new drug. It is therefore usual
282 in NI studies to choose a smaller margin (M_2) that reflects the largest loss of effect that
283 would be clinically acceptable. This can be described as an absolute difference in effect
284 (typical of antibiotic trials) or as a fraction of the risk reduction provided by the control
285 (typical in cardiovascular outcome trials). Note that the clinically acceptable margin could
286 be relaxed if the test drug were shown to have some important advantage (e.g., on safety or
287 on a secondary endpoint).

288

289 The definitions used to describe these two versions of M are:

290

291 M_1 = the entire effect of the active control assumed to be present in the NI study
292 M_2 = the largest clinically acceptable difference (degree of inferiority) of the test drug
293 compared to the active control

294

295 M_1 is based on (1) the treatment effect estimated from the historical experience with the
296 active control drug, (2) assessment of the likelihood that the current effect of the active
297 control is similar to the past effect (the constancy assumption), and (3) assessment of the
298 quality of the NI trial, particularly looking for defects that could reduce a difference between

1086 conservative estimate, such as the pharmacologic similarity of the test and control drugs and
1087 pharmacodynamic effects of the new drug, rather than reflecting “automatic” discounting.
1088 Having considered these matters, if significant uncertainties remain, an approach that further
1089 discounts or reduces, say by 25%, the magnitude of the active control effect based on
1090 HESDE can be considered.

1091

1092 A closely related issue is adjustment of M_1 to reflect a finding that the population in the NI
1093 study was different from the historical study in such a way that what the historical experience
1094 shows would lead to a smaller effect size (e.g., a finding of a smaller effect in women would
1095 need to be considered in assessing the validity of M_1 if the NI study had substantially more
1096 women than the historical studies). In general, the assessment of the historical data should
1097 identify such differences so that plans for the NI study take this into account or so that the
1098 value of M_1 can be revisited in light of the study population included in the NI study.

1099

1100 C. Statistical Methods for NI Analysis

1101

1102 Several approaches are used to demonstrate statistically that the NI objective is met. Each
1103 statistical approach to demonstrating NI depends upon a number of factors including:

1104

- 1105 • What assumptions are made and how verifiable or empirically demonstrable these
1106 assumptions are
- 1107 • The degree to which judgment, both statistical and clinical, is exercised in accounting
1108 for the various uncertainties in the data from the current NI study and also from the
1109 clinical trials of the active control that are the basis for estimating its effect
- 1110 • The clinical judgment of how much of the treatment effect of the active comparator
1111 can be lost (M_2 selection)

1112

1113 As noted earlier, the two main approaches to demonstrating non-inferiority are the fixed
1114 margin method and the synthesis method.

1115

1116 Each of these statistical approaches uses the same data from the previously conducted
1117 controlled trials of the active control and the same data from the current NI study, but the
1118 approaches are different in several ways. The first is with regard to their emphasis on the
1119 specific determination for M_1 before determining M_2 . There is also a difference between
1120 them in how the data from the historical studies and the NI study are used or combined.

1121 What follows is a guide to the differences between the two approaches. Examples 1(A) and
1122 1(B) in the Appendix provide more detailed illustrations of how each of these approaches is
1123 used and interpreted. In general, the fixed margin approach is more conservative and treats
1124 the variance of the NI study and historical evidence distinctly. That is, a very large historical
1125 database will give a narrower CI and larger 95% lower bound for M_1 , but it will not directly
1126 figure into the test drug versus placebo calculation, as is done in the synthesis method.

1127 Concern about using the synthesis approach reflects our view that the method incorporates
1128 too much certainty about the past results into the NI comparison. We believe the fixed
1129 margin approach is preferable for ensuring that the test drug has an effect greater than

Contains Nonbinding Recommendations

Draft – Not for Implementation

1130 placebo (i.e., the NI margin M_1 is ruled out). However, the synthesis approach, appropriately
1131 conducted, can be considered in ruling out the clinical margin M_2 .

1132

1133 1. *The Fixed Margin Approach for Analysis of the NI Study*

1134

1135 Sections IV.B.2.a and B.2.b contain discussions of the basic statistical approach to estimating
1136 the active comparator treatment effect size from past controlled trials. The goal of these
1137 analyses is to define the margin M_1 , a fixed value, based on the past effect of the active
1138 control, which is intended to be no larger than the effect the active control is expected to have
1139 in the NI study. Whether M_1 is based on a single study or multiple studies, the observed (if
1140 there were multiple studies) or anticipated (if there is only one study) statistical variation of
1141 the treatment effect size should contribute to the ultimate choice of M_1 , as should any
1142 concerns about constancy. The selection of M_2 is then based on clinical judgment regarding
1143 how much of the M_1 active comparator treatment effect can be lost. The exercise of clinical
1144 judgment for the determination of M_2 should be applied after the determination of M_1 has
1145 been made based on the historical data and subsequent analysis.

1146

1147 All relevant studies of the active comparator and all randomized patients within these studies
1148 should generally be used in determining the margin M_1 because that provides a more reliable
1149 and, possibly, conservative estimate. The actual selection of which studies are used in a
1150 meta-analysis and how that selection is made can be complex and itself subject to judgment.
1151 See Examples 1(A), 3, and 4 that illustrate these points in the Appendix.

1152

1153 The design and analysis of the NI study, and its analysis using the fixed margin approach, is
1154 well known and described in ICH E9, section 3.3.2. This statistical approach relies upon the
1155 choice of a fixed non-inferiority margin that is pre-specified and part of the NI design. There
1156 is very little, however, in ICH E9 or ICH E10 that discusses just how to determine the
1157 margin. Although the constancy assumption and study quality issues are recognized, there is
1158 little discussion about how to adjust the margin because of such statistical or study data
1159 uncertainties. Any discounting of the historical evidence of the effect of the active control
1160 based on uncertainty of the constancy of the effect (e.g., because of changes in practice or
1161 concomitant treatment), which is an attempt to improve the estimate of the control effect in
1162 the NI study, affects the M_2 as well, as in most cases M_2 is a fraction of M_1 . M_2 might not be
1163 affected when it is very small compared to M_1 , as is the case in considering very effective
1164 drugs. It is critical to note that M_2 is a judgment that is made after M_1 is chosen, but M_2 , of
1165 course, can never be larger than M_1 . It is perhaps tempting to make up for uncertainty in M_1
1166 by demanding assurance of preservation of a larger fraction of M_1 by ruling out a smaller
1167 loss of effect (i.e., using a smaller M_2), but the temptation should be avoided. The first and
1168 most critical task in designing an NI study is obtaining the best estimate of the effect of the
1169 active control in the NI study (i.e., M_1).

1170

1171 Operationally, the fixed margin approach usually proceeds in the following manner. The
1172 active comparator effect size is calculated from past placebo-controlled studies. The lower
1173 bound of the confidence interval describing the effect of the active control in past studies, a
1174 single number, is selected as a conservative choice for the active comparator effect size.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1175 While traditionally the 95% confidence interval is used, there can be flexibility in this choice,
1176 such as a 90% confidence interval or even narrower, when the circumstances are appropriate
1177 to do so (e.g., strong evidence of a class effect, strong biomarker data). It is recognized that
1178 use of a fixed margin to define the control response is conservative as it picks a “worst case”
1179 out of a confidence interval that consists of values of effect that are all larger. This choice,
1180 however, is one response to the inherent uncertainty of estimates based on past studies,
1181 including the variability of those past estimates, and the possibility that changes in medical
1182 practice, or hard to recognize differences between the past studies and the current NI study,
1183 have made the past effect an overestimate of the active control effect in the new study.

1184
1185 Although some of the uncertainty about applicability of past results to the present is reflected
1186 in a conservative choice of margin (95% of CI lower bound) used to initiate consideration of
1187 M_1 , there may be further concerns about past variability and constancy that lead to a
1188 determination to discount this lower bound further in choosing M_1 to account for any sources
1189 of uncertainty and dissimilarities between the historical data and the NI study to be
1190 conducted, as discussed in the earlier sections. Following this, a clinical judgment is made as
1191 to how much of this effect should be preserved. This clinical judgment could choose M_2 to
1192 be the same as M_1 , but as noted, where the treatment effect is important (e.g., an effect on
1193 mortality) it is usual to ask that a reasonable fraction of the control effect be preserved, by
1194 making M_2 , the loss of effect to be ruled out, smaller than M_1 . Choosing M_2 as 50% of M_1
1195 has become usual practice for cardiovascular (CV) outcome studies, whereas in antibiotic
1196 trials, where effect sizes are relatively large, a 10-15% NI margin for M_2 is common. Note
1197 that the M_2 of 50% of M_1 is on a relative scale, whereas the 10-15% is on the absolute scale
1198 for antibiotic drugs. The analysis of the NI study involves only the data from the NI study,
1199 and the test of the hypothesis that inferiority greater than the M_2 margin has been excluded is
1200 statistically similar to showing that the 95% CI in a superiority study excludes a difference of
1201 zero.

1202
1203 Thus, there are two confidence intervals involved in the fixed margin approach, one from the
1204 historical data, where one uses the lower bound to choose M_1 , and one from the NI study (to
1205 rule out $C-T > M_2$); in this example both intervals are 95% confidence intervals. That is why
1206 this fixed margin approach is sometimes called the 95%-95% method. It should be
1207 appreciated that the analysis of the NI study (ruling out a difference $> M_2$ by examining the
1208 lower bound of the CI for C-T) is the analysis that is based on the randomized comparison in
1209 the NI study, in contrast to the determination of M_1 , which is not based on a concurrent
1210 randomization.

1211
1212 Separating the process of estimating the treatment effect of the active comparator based upon
1213 the historical data (i.e., choice of M_1) from the analysis of the NI study has some advantages
1214 and disadvantages. Two important advantages are that it provides a single number that is
1215 clinically understandable for an M_1 (and derived M_2) and that it provides a basis for planning
1216 the sample size of the NI study to achieve statistical control of Type 1 error and the power
1217 needed for the NI study to meet its objective for the pre-specified NI margin. One arguable
1218 disadvantage is that the method is statistically not efficient because it uses the two confidence
1219 interval approach rather than a combined estimate of the statistical variability of the historical

Contains Nonbinding Recommendations

Draft – Not for Implementation

1220 and NI study data. Nevertheless, use of the fixed margin is readily understood, particularly
1221 by non-statisticians, and is only somewhat conservative compared to an analysis using the
1222 synthesis approach. Decisions to discount the M_1 further or, where appropriate, to use a
1223 narrower confidence interval, are easily explained, and can make the fixed margin approach
1224 more or less conservative.

1225

1226 Deciding on the NI clinical margin M_2 is also a relatively straightforward concept. It is
1227 plainly a matter of judgment about how much of the treatment effect must be shown to be
1228 preserved, a consideration that may reflect the seriousness of the outcome, the benefit of the
1229 active comparator, and the relative safety profiles of the test and comparator. It also has
1230 major practical implications. In large cardiovascular studies, it is unusual to seek retention of
1231 more than 50% of the control drug effect even if this might be clinically reasonable, because
1232 doing so will usually make the study size infeasible.

1233

1234 The fixed margin approach considers the NI margin as a single number, fixed in advance of
1235 the NI study. The hypothesis tested in the NI study determines whether the comparison of
1236 the test drug to the active control meets the specified NI criterion, assuming, of course, that
1237 the active control had at least its expected effect (equal to M_1) and that the study therefore
1238 had assay sensitivity. A successful NI conclusion, ruling out a difference $> M_1$, shows that
1239 the test drug is effective (just as a superiority study showing a significant effect at $p \leq 0.05$
1240 does) and, if a difference $> M_2$ is also ruled out, shows that the new drug preserves the
1241 desired fraction of the control drug's effect. This statistical test of hypothesis is not formally
1242 directed at determining whether the test drug would have been superior to a placebo, had a
1243 placebo group been included in the NI study, but it leads to a similar conclusion by ruling out
1244 the possibility that the test drug is inferior to the control by more than an amount equal to the
1245 whole effect of the control compared to placebo (that effect being known from past studies).

1246

1247 The possible outcomes of such trials are shown in Figures 2 and 3 in section III of this
1248 guidance.

1249

1250 *2. The Synthesis Approach for Analysis of NI*

1251

1252 An alternative statistical approach is known as the synthesis approach because it combines or
1253 synthesizes the data from the historical trials and the current NI trial, reflecting the variability
1254 in the two data sets (the current NI study and the past studies used to determine HESDE).

1255 The synthesis method is designed to directly address the question of whether the test product
1256 would have been superior to a placebo had a placebo been in the NI study, and also to
1257 address the related question of what fraction of the active comparator's effect is maintained
1258 (the loss to be ruled out) by the test product. In the synthesis approach, the NI margin is not
1259 predetermined, but the outcome of the NI study, a consideration of the effect of the test agent
1260 vs. placebo, can be judged for adequacy.

1261

1262 Although the synthesis approach combines the data from the historical trials into the
1263 comparison of the concurrent active comparator and the test drug in the NI study, a direct
1264 randomized concurrent comparison with a placebo is of course not possible, as the placebo

1265 group is not a concurrent control and there is no randomization to such a group within the NI
1266 study. The imputed comparison with a placebo group that is not in the NI study thus rests on
1267 the validity of several assumptions, just as the fixed margin approach does. The critical
1268 assumption of the constancy of the active control effect size derived from the historical
1269 controlled trials is just as important when the synthesis method is used.

1270

1271 Because of the way the variance of the historical data and the NI data are combined for the
1272 synthesis test, the synthesis test is more efficient (uses a smaller sample size or achieves
1273 greater power for the same sample size) than the fixed margin approach but requires
1274 assumptions that may not be appropriate. The statistical efficiency of the synthesis approach
1275 derives primarily from how the standard error of the comparison of test product to active
1276 comparator is dealt with. See Appendix, Example 1(B), for a comparison of the two methods
1277 and the variance calculations.

1278

1279 The synthesis approach does not specify a fixed NI margin. Rather, the method combines (or
1280 synthesizes) the estimate of treatment effect relative to the control from the NI trial with the
1281 estimate of the control effect from a meta-analysis of historical trials. The method treats both
1282 sources of data as if they came from the same randomized trial, to project where the placebo
1283 effect would have been had the placebo been present in the NI trial. The synthesis process
1284 makes use of the variability from the NI trial and the historical trials and yields one
1285 confidence interval for testing the NI hypothesis that the treatment preserves a fixed fraction
1286 of the control effect, without actually specifying that control effect or a specific fixed NI
1287 margin based on the control effect. Clinical judgment is used to pre-specify an acceptable
1288 fraction of the control therapy's effect that should be retained by the test drug, regardless of
1289 the magnitude of the control effect.

1290

1291 A disadvantage of the synthesis approach, however, is that it does not allow for a pre-
1292 specification of the actual size or magnitude of the NI margin M_1 , so the clinical judgment to
1293 determine the choice of M_2 is difficult and is generally not made until results are seen.
1294 Moreover, it may be unrealistic to assign the same weight to the variance of the historical
1295 outcome data and to that of the concurrent randomized NI treatment. As also noted, the
1296 efficiency of the fixed margin approach can sometimes be enhanced either formally, by
1297 including more trials (e.g., of related drugs) in the historical meta-analysis, and thereby
1298 increasing the margin M_1 , or, as a matter of judgment, by considering pharmacologic
1299 similarities between the control and test drugs, effects on pertinent biomarkers (e.g., tumor
1300 response rate), all of which could lead to choice of a fixed margin based on a less extreme
1301 boundary of the confidence interval (e.g., 80% instead of 95%).

1302

1303 **D. Considerations for Selecting M_2 , the Clinical Margin, and the Role of** 1304 **Subjective Judgment**

1305

1306 M_2 is the margin that is the pre-specified NI margin that should be met in an NI study. The
1307 determination of M_2 is based on clinical judgment and is usually calculated by taking a
1308 percentage or fraction of M_1 . The clinical judgment in determining M_2 may take into account
1309 the actual disease incidence or prevalence and its impact on the practicality of sample sizes

1310 that would have to be accrued for a study. There can be flexibility in the M_2 margin, for
1311 example, when:

1312

- 1313 (1) The difference between the active comparator response rate and the spontaneous
1314 response rate is large;
- 1315 (2) The primary endpoint does not involve an irreversible outcome such as death (in
1316 general, the M_2 margin will be more stringent when treatment failure results in an
1317 irreversible outcome);
- 1318 (3) The test product is associated with fewer serious adverse effects than other therapies
1319 already available;
- 1320 (4) The test product is in a new pharmacologic category and has been shown to be
1321 tolerated by patients who do not tolerate therapies that are already available.

1322

1323 There is also a difference in implication when the study NI conclusion is “not quite”
1324 significant (M_1 is not excluded) for M_1 and when this is the case for M_2 . Failure to exclude
1325 inferiority equal to M_1 means there is no assurance of any effect. Just as, for a placebo-
1326 controlled trial, it would be most unusual to accept as positive a study with $p > 0.05$, it would
1327 be most unusual to accept an NI study where the upper bound of 95% CI was $> M_1$. On the
1328 other hand, failing to exclude M_2 by a small amount means that instead of ruling out a loss of
1329 50% of M_1 , you have ruled out, say, a 48% loss, not necessarily a definitive failure. As noted
1330 above, we would also consider the less conservative synthesis approach in assessing M_2 .

1331

1332 **E. Estimating the Sample Size for an NI Study**

1333

1334 It is important to plan the sample size for an NI clinical trial so that the trial will have the
1335 statistical power to conclude that the NI margin is ruled out if the test drug is truly non-
1336 inferior. This is not always an easy task. At the protocol planning stage, using the fixed
1337 margin approach, the magnitude of the NI margin will be specified; the sample size must be
1338 based on the need to rule out inferiority greater than M_2 . This should usually be based on an
1339 NI using a fixed margin approach. The margin to be ruled out is the most critical component
1340 of the sample size planning, but the variance of the estimate of the treatment effects will not
1341 be known and it is also critical. A further problem is posed by the possibility that event rates
1342 will be lower in the new study. In this case, if the NI margin is expressed as, for example,
1343 ruling out (at the upper bound of the 95% CI for C-T) an increase in risk of 25%, this will be
1344 far easier when the event rate on active control is 8% than when it is 4%, even if the active
1345 control is superior to placebo by the same absolute 20% difference. This problem is not
1346 different from specifying sample size in a superiority trial. It too depends on the event rate,
1347 and it is common to examine blinded data during the trial to see if the event rate is
1348 unexpectedly low. A similar approach could be applied in an NI trial with upward
1349 adjustment of the sample size if the event rate is unexpectedly low. There is one further
1350 consideration. If, in reality, the test drug is somewhat more effective than the control, it will
1351 be easier to rule out any given NI margin and a smaller sample size could be used. A
1352 somewhat less effective test drug will, of course, require a larger sample size.

1353

1354 **F. Potential Biases in an NI Study**

1355
1356 Traditionally, analysis of the results of randomized clinical superiority trials follows the
1357 intent-to-treat principle, namely, that all randomized patients are analyzed according to the
1358 treatment to which they were randomized. This analysis is intended to avoid various biases
1359 associated with patients switching treatment, selection bias, and dropout/withdrawal patterns
1360 that may confound the observed treatment effect. This is recognized as a potentially
1361 conservative analysis. Including patient outcomes that occur after a patient has stopped the
1362 treatment, for example, or show poor compliance with treatment, would be expected to bias
1363 the analysis toward the null (no treatment difference). Intent-to-treat (ITT) analyses in
1364 superiority trials are nonetheless preferred because they protect against the kinds of bias that
1365 might be associated with early departure from the study. In non-inferiority trials, many kinds
1366 of problems fatal to a superiority trial, such as non-adherence, misclassification of the
1367 primary endpoint, or measurement problems more generally (i.e., “noise”), or many dropouts
1368 who must be assessed as part of the treated group, can bias toward no treatment difference
1369 (success) and undermine the validity of the trial, creating apparent non-inferiority where it
1370 did not really exist. Although an “as-treated” analysis is therefore often suggested as the
1371 primary analysis for NI studies, there are also significant concerns with the possibility of
1372 informative censoring in an as-treated analysis. It is therefore important to conduct both ITT
1373 and as-treated analyses in NI studies. Differences in results using the two analyses will need
1374 close examination. The best advice for conducting an NI study is to be aware at the planning
1375 stage of these potential issues and to monitor the trial in a manner that minimizes these
1376 problems, as they can seriously affect the validity of an NI study.

1377
1378 Other sources of bias that could occur in any study are also of concern in the NI study and
1379 are of particular concern in an open label study. For such open label NI studies, how best to
1380 ensure unbiased assessment of endpoints, unbiased decisions about inclusion of patients in
1381 the analysis, and a wide variety of other potential biases, need particular attention.

1382
1383 **G. Role of Adaptive Designs in NI Studies — Sample Size Re-estimation to**
1384 **Increase the Size of an NI Trial**

1385
1386 Because it may be difficult to adequately plan the sample size for any study, including an NI
1387 study, especially when assumptions like the event rate may change from the planning phase
1388 to the study conduct, adaptive study designs that can allow for the prospective re-estimation
1389 of a larger sample size can be considered. The most critical single consideration in such
1390 designs is precise knowledge about whether there is unblinding as to treatment. Sample size
1391 re-estimation, if based on a blinded analysis of the overall variance estimate or the overall
1392 event rate, without knowledge of or a comparison of the unblinded treatment group response
1393 rates or the differences between treatment groups, is not only acceptable but generally
1394 advisable. It is critical to provide reassurance and procedures that ensure maintenance of
1395 blinding.

1396
1397 If an adaptive design that allows unblinding is contemplated, then the design features and
1398 procedures for protection of the integrity of the trial need to be clearly stated in the protocol

1399 for the trial. Some adaptive designs may include an independent Data Monitoring
1400 Committee (DMC) to monitor the planned adaptation. The DMC charter should address
1401 procedures for the sharing and blinding of data, and the procedures used to maintain a
1402 firewall between those who do, and those who do not view unblinded data. Some of these
1403 issues will be addressed in a companion guidance on Adaptive Study Designs.

1404

1405 **H. Testing NI and Superiority in an NI Study**

1406

1407 In general, when there is only one endpoint and one dose of the test treatment, a planned NI
1408 study can be tested for superiority without a need for Type 1 error alpha correction. That is,
1409 the same 95% or higher confidence interval employed for testing non-inferiority with the pre-
1410 specified fixed margin can be used to test superiority. One can also think of this as a two-
1411 stage analysis in which the showing of NI using a 95% confidence interval (invariably
1412 successful if the test drug is actually superior), is then followed sequentially by superiority
1413 testing. This sequential testing has the Type I error rates for both non-inferiority and
1414 superiority controlled at a level of no more than 5%. A non-inferiority showing after a failed
1415 superiority study, in contrast, gives a generally uncertain result, and such a study would
1416 generally be considered a failed study. Thus, successful showing of non-inferiority allows
1417 superiority testing but a failed showing of superiority would yield credible evidence of non-
1418 inferiority only if the study were designed as a non-inferiority study (e.g., the NI margin must
1419 be pre-specified, and assay sensitivity and HESDE must be established).

1420

1421 When there are multiple endpoints or multiple doses of the test treatment evaluated in an NI
1422 study, the valid statistical decision tree can be very complex. Using the same 95%
1423 confidence interval to test non-inferiority and superiority at each endpoint level or at each
1424 dose may inflate the overall Type I error rate associated with drawing one or more false
1425 conclusions from such multiple comparisons, regardless of whether they are non-inferiority
1426 or superiority testing. Thus, for any statistical decision tree composed of tests of superiority
1427 and non-inferiority in multiple comparison settings, it is imperative to evaluate the overall
1428 Type I error rate for all the comparisons involved in the testing and make appropriate
1429 statistical adjustments.

1430

1431 Some of the problems in interpreting the results of non-inferiority analyses are more subtle
1432 than those with superiority testing. In particular, as noted previously, design or conduct
1433 problems such as medication non-compliance or misclassification/measurement error, errors
1434 that would be fatal to success in a superiority study, can lead to apparently favorable (results)
1435 in a non-inferiority study.

1436

1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481

V. COMMONLY ASKED QUESTIONS AND GENERAL GUIDANCE

1. Can a margin be defined when there are no placebo-controlled trials for the active control for the disease being assessed?

If the active control has shown superiority to other active treatments in the past, the difference demonstrated represents a conservative estimate of HESDE, one that can certainly serve as a basis for choosing M_1 . It may also be possible that trials of the active control in related diseases are relevant. The more difficult question is whether historical experience from nonconcurrently controlled trials can be used to define the NI margin. The answer is that it can, but the circumstances are similar to those in which a historically controlled trial can be persuasive (see ICH E-10). First, there should be a good estimate of the historical spontaneous cure rate or outcome without treatment. Examination of medical literature and other sources of information may provide data upon which to base these estimates (e.g., historical information on natural history or the results of ineffective therapy). Second, the cure rate of the active control should be estimated from historical experience, preferably from multiple experiences in various settings, and should be substantially different from the untreated rate. For example, if the spontaneous cure rate of a disease is 10-20% and the cure rate with an active control is 70-80%, these are substantially different and an acceptable margin, generally chosen conservatively, can probably be identified for M_1 . The clinically acceptable loss of this effect can then be determined for M_2 . Estimates of the cure rate of the active control should be based upon data from clinical trials, even if these are not controlled, and it is critical to be sure the trial patients and untreated patients are similarly defined and selected. Example 2 in the Appendix illustrates a case of this kind, in which it was concluded that a margin could be defined despite the absence of placebo-controlled trials of the active control. It becomes more difficult to identify a margin when the difference between the spontaneous cure rate and active drug cure rate is smaller. For example, if the historical spontaneous cure rate is 40% and the active control rate is 55%, it would not be credible to identify the NI margin in this case as 15%, as such a small difference could easily be the result of different disease definition or ancillary therapy. When the historical cure rates for the active control and the cure rate in patients who receive no treatment are not known at all from actual studies (i.e., are just based on clinical impressions), it will be difficult or impossible to define an NI margin.

2. Can the margin M_2 be flexible?

As indicated in sections III and IV, there is a critical difference between demonstrating in the NI study that the margins M_1 and M_2 have been met. M_1 is used to determine whether the NI study shows that the test drug has any effect at all. Accepting a result in which the 95% CI did not rule out loss of M_1 would be similar to accepting, as showing effectiveness, a superiority study whose estimated treatment effect was not significant at $p \leq 0.05$. M_2 , in contrast, represents a clinical judgment about what level of loss of the active control effect is acceptable. A typical value for M_2 is often 50% of M_1 , at least

1482 partly because the sample sizes needed to rule out a smaller loss become impractically
1483 large. In this case, there is a better argument for some degree of flexibility if the study
1484 did not quite rule out the M_2 margin; there might be reason to consider, for example,
1485 assurance of 48% retention (but not the expected 50%) for M_2 as acceptable. We have
1486 also concluded that the fixed margin method, more conservative but with fewer
1487 assumptions, should generally be used in ensuring that loss of M_1 is ruled out but that the
1488 synthesis method can be used to assess M_2 . Of course, allowing too much inferiority of
1489 the test drug to the standard, especially for endpoints of mortality and serious morbidity,
1490 would clearly not be acceptable.

1491

1492 **3. Can prior information or other data (e.g., studies of related drugs, pharmacologic**
1493 **effects) be considered statistically in choosing the NI margins or in deciding whether**
1494 **the NI study has demonstrated its objective?**

1495

1496 Prior information could be characterized in a statistical model or in a Bayesian
1497 framework by taking into account such factors as evidence of effects in multiple related
1498 indications or on many endpoints. Such information might be used in determining M_1 in
1499 a more flexible (less conservative) manner. For example, if multiple studies provide very
1500 homogeneous results for one or more important endpoints it may be possible to use the
1501 90% lower bound rather than the 95% lower bound of the CI to determine the active
1502 control effect size. Similarly, if there were additional supporting evidence for the clinical
1503 effect of the test drug, such as prior information on the efficacy of the test drug in related
1504 diseases or in a compelling animal model, or an effect on an important biomarker (e.g.,
1505 tumor response rate), or evidence that pharmacologically related drugs were clearly
1506 effective in the condition being studied, such prior information would increase the
1507 evidence for the plausibility of the intended NI effect of the test drug, which might allow
1508 use of a less conservative estimate of effect than the 95% lower bound of the confidence
1509 interval for C-T in the NI study. Finally, a statistical model such as a regression
1510 adjustment may be applied to the NI study analysis if the covariates for patients in the
1511 historical clinical studies are distributed differently from those of patients in the current
1512 NI study. This adjustment may, in some situations, reduce the variance of the NI test and
1513 increase the ability of the comparison to meet the NI margin. In other situations, where
1514 there is more heterogeneity of the covariates, the variance may be increased, adversely
1515 impacting the comparison.

1516

1517 **4. Can a drug product be used as the active comparator in a study designed to show**
1518 **non-inferiority if its labeling does not have the indication for the disease being**
1519 **studied, and could published reports in the literature be used to support a treatment**
1520 **effect of the active control?**

1521

1522 The active control does not have to be labeled for the indication being studied in the NI
1523 study, as long as there are adequate data to support the chosen NI margin. FDA does, in
1524 some cases, rely on published literature and has done so in carrying out the meta-analyses
1525 of the active control used to define NI margins. An FDA guidance for industry on
1526 *Providing Clinical Evidence of Effectiveness for Human Drug and Biological Products*

1527 describes the approach to considering the use of literature in providing evidence of
1528 effectiveness, and similar considerations would apply here. Among these considerations
1529 are the quality of the publications (the level of detail provided), the difficulty of assessing
1530 the endpoints used, changes in practice between the present and the time of the studies,
1531 whether FDA has reviewed some or all of the studies, and whether FDA and the sponsor
1532 have access to the original data. As noted above, the endpoint for the NI study could be
1533 different (e.g., death, heart attack, and stroke) from the primary endpoint (cardiovascular
1534 death) in the studies if the alternative endpoint is well assessed (see also question 6).
1535

1536 **5. If the active control drug is approved for the indication that is being studied, does**
1537 **the margin need to be justified, or if the active control drug has been used as an**
1538 **active comparator in the past in another study of design similar to the current study**
1539 **and a margin has been justified previously, can one simply refer to the previous**
1540 **margin used?**

1541
1542 When an active control drug is approved, the effect size for the indication is not usually
1543 identified in a pooled analysis, nor is the variability of that effect size in the various trials
1544 calculated. It would therefore be difficult to base the NI margin on the label of the active
1545 control drug. On the other hand, FDA’s reliance on the studies for approval would
1546 support the view that the quality of the studies was acceptable and that the studies could
1547 contribute to a determination of the NI margin. In general, approval of a drug is based on
1548 showing superiority to placebo, usually in at least two studies, but FDA may not have
1549 critically assessed effect size and may not have closely analyzed “failed” studies. In
1550 general, FDA will usually not have carried out a meta-analysis of the trials. It is therefore
1551 essential to use the data from all available controlled trials (unless a trial has a significant
1552 defect), including trials conducted after marketing, to calculate a reasonable estimate of
1553 the actual control effect size, as described above. If the active-control data have been
1554 used to define a NI margin for another study, it is important to determine that the
1555 previous conclusion is applicable to the new study, but in general such prior use should
1556 indicate that FDA has assessed the NI margin for a NI study with similar endpoints and
1557 population.
1558

1559 **6. What are the choices of endpoints to be aware of before designing a non-inferiority**
1560 **trial design?**

1561
1562 The endpoints chosen for clinical trials (superiority or NI) reflect the event rate in the
1563 population, the importance of the event, and practical considerations, notably whether the
1564 event rates will allow a study of reasonable size. In NI studies, the endpoint must be one
1565 for which there is a good basis for knowing the effect of the active control. The endpoint
1566 used need not necessarily be the endpoint used in the historical trials or the effectiveness
1567 endpoint claimed in labeling. Past trials, for example, with mortality endpoints could, if
1568 data were available, be the basis for estimating an effect on a composite endpoint
1569 (cardiovascular mortality, myocardial infarction, and stroke), if that were the desired
1570 endpoint for the NI study. Such a change might be sought because it would permit a
1571 smaller study or was more feasible given current event rates.

1572

1573

1574

7. Are there circumstances where it may not be feasible to perform an NI study?

1575

Unfortunately, these are many, including some where a placebo-controlled study would not be considered ethical. Some examples include the following:

1576

1577

1578

- The treatment effect may be so small that the sample size required to do a non-inferiority study may not be feasible.

1579

1580

- There is large study-to-study variability in the treatment effect. In this case, the treatment effect may not be sufficiently reproducible to allow for the determination of a sufficiently reliable estimate of M_1 .

1581

1582

1583

- There is no historical evidence to determine a non-inferiority margin.

1584

- Medical practice has changed so much (e.g., the active control is always used with additional drugs) that the effect of the active control in the historical studies is not clearly relevant to the current study.

1585

1586

1587

1588

8. In a situation where a placebo-controlled trial would be considered unethical, but a non-inferiority study cannot be performed, what are the options?

1589

1590

In that case it may be possible to design a superiority study that would be considered ethical. These possibilities are discussed in section III of this guidance and ICH E-10, and include the following:

1591

1592

1593

1594

- When the new drug and established treatment are pharmacologically distinct, an add-on study where the test drug and placebo are each added to the established treatment.
- A study in patients who do not respond to the established therapy. It may be possible to do a placebo-controlled trial in those patients. To establish specific effectiveness in non-responders, the study should randomize to test drug and the failed therapy and show superiority of the test drug.
- A study in patients who cannot tolerate the established effective therapy.
- A study of a population in which the effect of available therapy is not established.
- For a drug with dose-related side effects, and where a dose lower than the usual dose would be considered ethical, a dose-response study may be possible.

1595

1596

1597

1598

1599

1600

1601

1602

1603

1604

1605

9. When will a single NI study be sufficient to support effectiveness?

1606

1607

Several sections above touch on this question, notably III.B.2, which discusses it in detail. Briefly, reliance on a single study in the NI setting is based on considerations similar to reliance on a single study in the superiority setting, with the additional consideration of the stringency of showing NI using the M_2 NI margin. Many of these factors are described in the guidance for industry on *Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products*, and include prior supportive information, such as results with pharmacologically similar agents (a very common consideration, as the NI study will often compare drugs of the same pharmacologic class), support from credible biomarker information (tumor responses, ACE inhibition,

1608

1609

1610

1611

1612

1613

1614

1615

1616

Contains Nonbinding Recommendations

Draft – Not for Implementation

1617 beta blockade), and a statistically persuasive result. With respect to the latter, it is noted
1618 above that a finding of NI based on excluding a treatment difference $> M_2$ provides very
1619 strong evidence (generally equivalent to a $p < 0.001$ in a superiority setting) that the test
1620 treatment has an effect > 0 . For all these reasons, most NI studies with outcome
1621 endpoints, if clearly successful, will be supportive as single studies. Of course, the
1622 importance of the study endpoint will influence the level of assurance needed, in a single
1623 study or multiple studies, that no more than M_2 has been lost.

1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668

APPENDIX — EXAMPLES

The following five examples derived from publicly available information (see references following examples) illustrate different aspects of the process of choosing a NI margin, of the application of a method of NI analysis, and other considerations relevant to whether it is possible to conduct and interpret the results of a NI study

Example 1(A): Determination of an NI Margin for a New Anticoagulant — Fixed Margin Approach

This example will demonstrate the following points:

- The determination of the NI margin (M_1) using the fixed margin approach
- How to select and assess the randomized trials of the active control on which to base the estimate of active comparator treatment effect.
- How to assess whether the assumption of assay sensitivity is appropriate, and whether the constancy assumption is reasonable for this drug class.
- Why it is appropriate to use a conservative choice (e.g., 95% lower bound) for estimating the treatment effect size of the active comparator, accounting for between-study variability, and considering other uncertainties in the randomized trial data.
- The use of the lower bound of the 95% confidence interval in the NI study for C-T to demonstrate non-inferiority.

SPORTIF V is an NI study that tested the novel anticoagulant ximelagatran against the active control warfarin. Warfarin is a highly effective, orally active anticoagulant that is approved in the United States for the treatment of patients with non-valvular atrial fibrillation at risk of thromboembolic complications (e.g., stroke, TIA, etc.). There are six placebo-controlled studies of warfarin involving the treatment of patients with non-valvular atrial fibrillation, all published between the years 1989 and 1993. The primary results of these studies are summarized in Table 1 and provide the basis for choosing the NI margin for SPORTIF V.

The point estimate of the event rate on warfarin compared to placebo is favorable to warfarin in each of the 6 studies. The upper bound of the 95% confidence interval of the risk ratio calculated in each study is less than one in five of the six studies, indicating a statistically demonstrated treatment effect in each of these studies. The one exception is the CAFA study. However, this study was reportedly stopped early because of favorable results published from the AFASAK and SPAF I studies (Connolly et al. 1991). Although the CAFA study was stopped early, a step that can sometimes lead to an overestimate of effect, the data from this study appear relevant in characterizing the overall evidence of effectiveness of warfarin because there is no reason to think it was stopped for early success, introducing a possible favorable bias. These placebo controlled studies of warfarin in

Contains Nonbinding Recommendations

Draft – Not for Implementation

1702 subjects had a prior history of stroke or TIA. None of the other studies had such a
 1703 requirement. Subjects enrolled into the EAFT study were thus at higher risk than subjects in
 1704 the other studies, presumably leading to the higher event rates in both the warfarin and
 1705 placebo arms, shown in Table 1. The higher event rates in the EAFT study may also have
 1706 been influenced by the relatively long duration of follow-up or the fact that the primary
 1707 endpoint definition was broader, including vascular deaths and non-fatal myocardial
 1708 infarctions, which might have been less affected by coumadin, leading to a lower risk
 1709 reduction. This was not in fact seen. All in all, the results are quite consistent (with the
 1710 exception of CAFA), a relatively reassuring outcome.

1711

1712 **Table 2: Demographic Variables, Clinical Characteristics, and Endpoints of Warfarin AF Studies**

	AFASAK	BAATAF	CAFA	SPAF	VA	EAFT	SPORTIF V
Age years (mean)	73	69	68	65	67	71	72
Sex (%) Male	53%	75%	76%	74%	100%	59%	70%
h/o stroke or TIA (%)	6%	3%	3%	8%	0%	100%	18.3%
h/o HTN (%)	32%	51%	43%	49%	55%	43%	81%
≥65 years old & CAD (%)*	8%	10-16%	12-15%	7%	17%	7%	41%
>65 years old & DM (%)*	7-10%	14-16%	10-14%	13%	17%	12%	19%
h/o LV dysfunction (%)*	50%	24-28%	20-23%	9%	31%	8%	39%
Mean BP at BL (mm Hg)	NA	NA	NA	130/78	NA	145/84	133/77
Target INR	2.8-4.2	1.5-2.7	2-3	2-4.5	1.4-2.8	2.5-4.0	2-3
Primary endpoint	Stroke, TIA, systemic embolism	Ischemic stroke	Ischemic stroke and systemic embolism	Ischemic stroke and systemic embolism	Ischemic stroke	Vascular death, NF MI, stroke, systemic embolism	Stroke (ischemic + hemorrhagic) and systemic embolism

1713 * = Not possible to verify whether definitions of CAD, DM, and LV dysfunction

1714 were the same in comparing the historic studies and SPORTIF V.

1715 NA = Not available

1716

1717 At the time the SPORTIF V study was reviewed, concerns about whether the constancy
 1718 assumption held and other factors led to the consideration of whether discounting of the
 1719 effect size would be appropriate (see discussion of discounting in section IV of this
 1720 guidance). We now believe the historic results are reasonably likely to be consistent with
 1721 results that would be seen today so that discounting was not necessary. To calculate M_1 , the
 1722 relative risks in each of the six studies were combined using a random effects model to give a
 1723 point estimate of 0.361 for the relative risk with a confidence interval of (0.248, 0.527). The
 1724 95% CI upper bound of 0.527 represents a 47% risk reduction, which translates into a risk
 1725 increase of about 90% from not being on warfarin ($1/0.527 = 1.898$) (i.e., what would be seen
 1726 if the test drug had no effect). Thus, M_1 (in terms of the hazard ratio favoring the control to
 1727 be ruled out) is 1.898.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1728

1729 It was considered clinically necessary to show that the test drug preserved a substantial
1730 fraction of the warfarin effect. The clinical margin M_2 representing the largest acceptable
1731 inferiority of the test to control, was therefore set at 50% of M_1 . As described in section IV
1732 of the guidance, we calculate M_2 , using the log hazard risk ratios, as 1.378, 95% CI for C-T <
1733 1.378.

1734

1735 In the SPORTIF V study, the point estimate of the relative risk was 1.39 and the two-sided
1736 95% confidence interval for the relative risk was (0.91, 2.12). Thus, in this example, the
1737 non-inferiority of ximelegatran to warfarin is not demonstrated because the upper limit (2.12)
1738 is greater than M_2 (=1.378). Indeed, it does not even demonstrate that M_1 (=1.898) has been
1739 excluded.

1740

1741 This example illustrates the fixed margin approach and what is often called the “two 95%
1742 confidence interval approach.” That is, a two-sided 95% confidence interval is used for the
1743 historical data to select M_1 , and a two-sided 95% confidence interval is used to test whether
1744 M_2 has been ruled out, similar to controlling the Type 1 error of the NI study at one-sided
1745 2.5%.

1746 **Example 1(B): Application of the Synthesis Method to the Above Example 1(A)**

1747

1748 This example demonstrates the following:

1749

1750 • The critical features of the synthesis approach to demonstrating the NI of a new
1751 anticoagulant.

1752

1753 • The calculations and sources of statistical variability that are incorporated in the
1754 synthesis approach.

1755

1756 • The main differences in interpretation of the fixed margin and the synthesis approaches
1757 when applied to the same set of studies and data.

1758

1759 In this example, we illustrate the synthesis method using the same data as Example 1(A),
1760 which consist of six studies comparing warfarin to placebo and one NI study comparing
1761 ximelegatran to warfarin. In contrast to the fixed margin method in Example 1(A), the
1762 synthesis method does not use a separate 95% confidence interval for this historical estimate
1763 of the effect of warfarin versus placebo and for the comparison in the NI study. Rather, the
1764 synthesis method is constructed to address the questions of whether ximelegatran preserves a
1765 specified percent, in this case 50% or one-half (versus placebo), of the effect of warfarin, and
1766 whether ximelegatran would be superior to a placebo, if one had been included as a
1767 randomized treatment group in the NI study. To accomplish this goal, the synthesis method
1768 makes a comparison of the effect of ximelegatran in the NI study to historical placebo data,
1769 an indirect comparison that is not based upon a randomized current placebo group. The
1770 synthesis method combines the data from the placebo-controlled studies of warfarin with the
1771 data from the NI study in such a way that a test of hypothesis is made to demonstrate that a
1772 certain percent of the effect of warfarin is retained in the NI study. A critical point
1773 distinguishing the synthesis method from the fixed margin method is that the M_1 effect size
1774 of warfarin is not specified in advance and is not required to be fixed prior to carrying out the
1775 synthesis method. But to carry out the analysis, an assumption needs to be made regarding
1776 the placebo comparison, namely, that the difference between control drug and placebo (had
1777 there been one) in the NI trial is the same as what was seen in the historical placebo-
1778 controlled trials of warfarin. The assumption is needed because there is no randomized
1779 comparison of warfarin and placebo in the NI trial. As a point of reference, we know from
1780 the previous example, 1(A), that the warfarin effect M_1 was estimated from the historical
1781 placebo studies to be a 47% risk reduction.

1782

1783 In this case, the synthesis method statistically tests the null hypothesis that the inferiority of
1784 ximelegatran compared to warfarin is less than 50% or one half of the risk reduction of
1785 warfarin compared to placebo, a question that the fixed margin method does not directly
1786 address because in the fixed margin method, the placebo is only present in the historical
1787 studies and not in the NI study. We carry out this test on the log relative risk scale, so that
1788 the null hypothesis can be written as:

Contains Nonbinding Recommendations

Draft – Not for Implementation

1789

1790

$H_0: \{\log\text{-Relative Risk of ximelegatran versus warfarin}\} \geq$

1791

$-\frac{1}{2} \{\log\text{-Mean Relative Risk of warfarin versus placebo}\}$

1792

A test of this hypothesis is performed by the expression below (the statistical test) that has the form of a quotient where the numerator is an estimate of the parameter defined in the null hypothesis by $\{\log\text{-Relative Risk of ximelegatran versus warfarin}\} + \frac{1}{2} \{\log\text{-Mean Relative Risk of warfarin versus placebo}\}$ and the denominator is an estimate of the standard error of the numerator. In this case, the estimated log-Relative Risk of ximelegatran versus warfarin is 0.329 (log of 1.39) with a standard error of 0.216 while the estimated log-Relative Risk of warfarin versus placebo is -1.02 (log of .527) with a standard error of 0.154. The estimate of the log warfarin effect is -1.02, and the standard error of this estimate is 0.154; these estimates are combined with the NI data as if all the data were in a randomized comparison with placebo. The synthesis test statistic is calculated as:

1793

1794

1795

1796

1797

1798

1799

1800

1801

1802

1803

$$\frac{0.329 + \frac{1}{2}\{-1.02\}}{\sqrt{0.216^2 + \left\{\frac{1}{2}\{0.154\}\right\}^2}} = -0.789$$

1804

1805

1806

1807

1808

1809

Assuming the statistic is normally distributed, it is then compared to -1.96 (for one-sided Type 1 error rate of 0.025). For this case, the value, -0.789, is not less (more negative) than -1.96, so we cannot reject the null hypothesis. Therefore, it cannot be concluded that an NI margin of 50% retention is satisfied.

1810

1811

1812

1813

1814

1815

1816

1817

1818

1819

To compare the fixed margin method with the synthesis method, recall that the fixed margin compares the upper or lower limits of two 95% confidence intervals, one for the NI study and one for the meta-analysis of the effect of warfarin. One might consider the fixed margin approach as conservative, as it compares to statistically “worst cases.” The synthesis method does not use two such worst cases. To provide a more detailed comparison of the approaches, the fixed margin approach can be expressed as using a test statistic similar to that of the synthesis approach.

The synthesis method concludes non-inferiority if

1820

$$\frac{0.329 + \frac{1}{2}\{-1.02\}}{\sqrt{0.216^2 + \left\{\frac{1}{2}\{0.154\}\right\}^2}} < -1.96$$

1821

Contains Nonbinding Recommendations

Draft – Not for Implementation

1822

1823 The fixed margin method concludes non-inferiority if

1824

1825
$$\frac{0.329 + \frac{1}{2}\{-1.02\}}{0.216 + \frac{1}{2}\{0.154\}} < -1.96$$

1826

1827 The critical difference between these two procedures is the form of the denominator, which
1828 expresses the standard errors of the expressions in the numerator. The synthesis standard
1829 error is always smaller than that of the fixed margin method when expressed in this manner.
1830 In most situations, the synthesis is therefore statistically more efficient (and would require a
1831 smaller sample size) than the fixed margin approach. Of course, the approach can be
1832 considered useful and valid only if the assumptions of the synthesis method can be
1833 considered satisfied. This is not always possible, generally because of concerns about
1834 constancy, that is, whether the historical differences from placebo would accurately describe
1835 the current differences from placebo.

1836

1837 The two procedures also cannot be directly compared because they have other differences
1838 that make their comparison problematic, notably the differences in how the statistical error
1839 rates, or Type 1 errors, are calculated and interpreted. The synthesis method, because of the
1840 way it makes the comparisons with a placebo, gives equal weight to the variance (or
1841 variability of the outcome data) in this historical estimate and the variance of the data
1842 obtained from the randomized comparison of the test drug and active comparator in the NI
1843 study. When the historical database is very large relative to the NI database, combining the
1844 historical data and NI together may suggest greater precision in the overall assessment of the
1845 NI study than is warranted given the fact that the placebo comparisons were from studies
1846 conducted in a different population, usually at a different time. In contrast, the fixed margin
1847 method controls a Type 1 error rate within the NI study that is conditioned on the pre-
1848 specified fixed NI margin, separately estimated from the historical active comparator data.
1849 The synthesis test method also does not estimate a fixed NI margin to be excluded (i.e., one
1850 depending only on the prior placebo-controlled data for the active comparator).

1851

1852 A general principle expressed in this guidance is the need to be conservative in the selection
1853 of the margin M_1 because that margin is critical to establishing that a test drug is effective in
1854 an NI study design. The M_1 margin is usually chosen conservatively because of the
1855 uncertainties associated with the validity of assumptions in an NI study and the reliance on
1856 historical active control comparisons. As noted, the fixed margin approach can be
1857 considered conservative in that several worst case situations (lower bounds of 95%
1858 confidence intervals) are used, one evaluating the historical evidence and another in the NI
1859 comparison. We recommend use of this conservative fixed margin approach to selecting the
1860 M_1 margin and to demonstrating in the NI study that the M_1 margin is excluded at the
1861 acceptable Type 1 error. The synthesis method, on the other hand, as described above, is less
1862 conservative. But this is reasonable, given that M_2 is considerably smaller (a more
1863 demanding margin) and that the presence of a control drug effect has been well established
1864 by ruling out loss of M_1 using the fixed margin approach. We therefore believe the NI study

Contains Nonbinding Recommendations

Draft – Not for Implementation

1865 should utilize a fixed margin approach to ruling out loss of M_1 but can use the synthesis
1866 method to establish that loss of effect greater than the clinically relevant margin M_2 has been
1867 ruled out.

1868 **Example 2: The Determination of a Non-Inferiority Margin for Complicated Urinary**
1869 **Tract Infection (cUTI) — Fixed Margin Approach**

1870
1871 This example will illustrate the following points:

- 1872
- 1873 • The use of the absolute difference in cure rates as the metric of treatment effect.
 - 1874 • The determination of a non-inferiority margin when there are no randomized active
1875 comparator placebo-controlled studies available for the indication of interest (in this
1876 case, cUTI).
 - 1877 • Estimating the placebo response rate in cUTI based upon data from uncomplicated
1878 urinary tract infections (a generally less severe form of urinary tract infection leading
1879 to a high, therefore conservative, estimate).
 - 1880 • The importance of seeking out all relevant studies for the margin determination and
1881 incorporating the limitations of the studies, the analyses, and the resulting estimates in
1882 the consideration of the resulting estimate of the non-inferiority margin.
 - 1883 • This approach (i.e., relying on data other than controlled trials of the active control) is
1884 credible only when the effect size is large, given its limitations.
- 1885

1886 The following steps were used to estimate the effectiveness of the active control.

- 1887
- 1888 1. Evaluation of the placebo response rate in uncomplicated urinary tract infection
1889 (uUTI)
 - 1890 2. Evaluation of outcomes in patients receiving inadequate or inappropriate therapy for
1891 complicated urinary tract infection (cUTI)/acute pyelonephritis (AP)
 - 1892 3. Evaluation of the active comparator's response rate (levofloxacin, in this case) for
1893 cUTI.
- 1894

1895 **Step 1: Placebo Response Rate for Uncomplicated Urinary Tract Infection (uUTI)**

1896
1897 Although there were no placebo-controlled complicated UTI studies available, three placebo-
1898 controlled studies in women with uncomplicated UTI were identified. Among these three
1899 studies there were differences in the duration of study drug, endpoints assessed, and the
1900 diagnostic criteria for significant bacteriuria. There were no placebo-controlled trials
1901 identified in men with UTI without significant co-morbid conditions, and the
1902 pathophysiology and natural history of UTI are different in men and women. It would be
1903 expected that placebo response rates would therefore be high in such studies compared to the
1904 untreated rate in cUTI and represent a conservative (high) estimate of the spontaneous cure
1905 rate in cUTI.

1906
1907 Microbiological eradication rate is generally used as the primary endpoint for UTI studies.
1908 In the three placebo-controlled studies identified for UTI, the bacteriological response rates
1909 were 95/227(42%) for the combined 8-10 and 35-49 days (Ferry et al.), 9/27(33%) at day 3
1910 (Christiaens et al.), and 8/18(44%) in 1 week (Dubi et al.). The bacteriologic criteria for
1911 entry used in the Ferry study were $\geq 10^3$ CFU/ml for primary pathogens, whereas $\geq 10^4$
1912 CFU/ml was used for the Christiaens study. Because a count of $\geq 10^5$ CFU/ml is more

1913 typically used as diagnostic criteria for a uropathogen, the studies could overestimate the
 1914 placebo response rates by including patients whose colony counts would not cause them to be
 1915 considered infected. The results are summarized in the following table.
 1916

Author	Type of UTI	Placebo	95% CI ¹
Ferry et al.	uUTI	95/227 (42%)	(35.4 %, 48.6%)
Christiaens et al.	Acute uUTI	9/27 (33%)	(16.5%, 54.0%)
Dubi et al.	uUTI	8/18 (44%)	(21.5%, 69.2%)

1917 ¹Exact Confidence Intervals

1918
 1919 Because of the unequal study population sizes, a weighted analysis is needed. The weighted
 1920 non-iterative method for random effects model using logit of the event rates described by
 1921 DerSimonian and Laird was used to obtain the estimate and its 95% CI; the weighted
 1922 estimate is 41.2% with 95% CI of (35.5%, 47.2%).
 1923

1924 **Step 2: Outcomes Subsequent to Inadequate or Inappropriate Antibacterial Therapy**
 1925 **for Complicated Urinary Tract Infection (cUTI)/AP**

1926
 1927 Three studies were identified in which some patients were treated with an antimicrobial drug
 1928 to which the bacteria causing their UTI were resistant (inadequate therapy). Eradication rates
 1929 for pathogens resistant to the antimicrobial drug may be considered as another way to
 1930 estimate the placebo effect in cUTI/AP. It should be noted, however, that the use of data
 1931 from inadequate therapy may result in an estimate that is higher than a true placebo, once
 1932 again a conservative estimate of effect, because even “inadequate” therapy may have some
 1933 effect on the patient’s infection.
 1934

Author	Type of UTI	Eradication Rates	95% CI ¹
Allais et al.	cUTI/AP	12/23 (52.2%)	(30.6%, 73.2%)
Fang et al.	cUTI/AP	4/28 (14.3%)	(4.0%, 32.7%)
Talan et al.	AP	7/14 (50.0%)	(23.0%, 77.0%)

1935 ¹Exact Confidence Intervals

1936
 1937 The data from the historical studies in Table 4 were combined to obtain a weighted estimate
 1938 of the inadequate therapy eradication rate and its corresponding two-sided 95% CI. The
 1939 weighted estimate using the DerSimonian and Laird approach (random effect model) is
 1940 36.8% with 95% CI of (15.4%, 64.9%).
 1941

1942 **Step 3: Active Comparator's Eradication Rate for Complicated UTI (cUTI)**

1943
 1944 To assess the eradication rates for the active comparator, levofloxacin, four cUTI studies
 1945 were considered, including two published studies and two studies submitted to the Agency
 1946 (Study A and Study B) that involved men and women ≥18 years old. The two studies from

1947 the medical literature had limitations. In the Peng study, the microbiological eradication rate
 1948 was evaluated on Day 5, while antibiotic therapy was still ongoing. This could have falsely
 1949 elevated the response rate. The Klimberg study was an open-label study, and was excluded
 1950 from the analysis because of concern about potential bias.

1951
 1952 The other two studies, Study A and Study B, were blinded controlled studies using
 1953 levofloxacin for the treatment of cUTI. In Study A, the microbiological eradication rate for
 1954 levofloxacin was 84.2% (154/183). In Study B, the microbiological eradication rate for
 1955 levofloxacin was 78.2% (252/321). The levofloxacin eradication rates for the Peng study and
 1956 Studies A and B are shown in Table 5. The weighted estimate of eradication rates using the
 1957 DerSimonian and Laird approach is 81.6% with 95% CI of (75.8%, 86.3%).

1958

Table 5: Historical Levofloxacin Data from Published cUTI Studies			
Author	Type of UTI	Levofloxacin Microbiological Eradication Rate	95% CI ¹
Peng et al.	cUTI	18/20 (90%)	(68.3%, 98.8%)
Study A	cUTI and AP	154/183 (84.2%)	(78.0%, 89.1%)
Study B	cUTI and AP	252/321 (78.2%)	(73.6%, 82.9%)

1959 ¹Exact confidence intervals

1960

1961 **Step 4: Estimated Non-Inferiority Margin for Complicated UTI (cUTI) Using**
 1962 **Levofloxacin as the Active Comparator**

1963

1964 The placebo eradication rate is estimated from the upper bound of the two-sided 95% CI for
 1965 the placebo eradication rate in uUTI (47%) and this estimate is supported by evidence based
 1966 on outcomes subsequent to inadequate or inappropriate therapy in cUTI (65%). The
 1967 estimated levofloxacin cure rate for sensitive organisms is 76% (using the lower bound of the
 1968 95% CI for the weighted levofloxacin response rate). Using the placebo eradication rate for
 1969 uUTI, the historical treatment effect can be calculated as 29% (=76%-47%). The treatment
 1970 effect based on outcomes following inadequate antibacterial therapy can be calculated as
 1971 11% (=76%-65%), providing supportive evidence.

1972

1973 **Major Limitations in This Example:**

1974

1975 Apart from the lack of a direct comparison of active control and placebo in cUTI, there were
 1976 various uncertainties in the historical estimates described above because of problems with
 1977 data quality, study design, population size, prognostic factors, and differences in the timing
 1978 of the microbiological endpoint assessments. On the other hand, the placebo eradication rate
 1979 was estimated based on placebo-controlled clinical studies assessing the antibacterial
 1980 treatment in a population (female subjects with uUTI) that would almost certainly give an
 1981 overestimate of the spontaneous or placebo eradication rate in cUTI, leading to a
 1982 conservative (low) estimate of the effect of the active control.

Contains Nonbinding Recommendations

Draft – Not for Implementation

1983

1984 **Discounting and Preservation of the Levofloxacin Treatment Effect:**

1985

1986 The various limitations and uncertainties in the historical data led to discounting of the
1987 calculated treatment effect of 29%. Thus, the active control treatment effect over placebo
1988 (M_1) was estimated as 14.5% based on a 50% discounting. For a serious illness, a substantial
1989 portion (at least 50% or more) of M_1 should be preserved. Accordingly, an NI margin of 7%
1990 was specified as M_2 based on clinical judgment.

1991

1992 **Example 3: Aspirin to Prevent Death or Death/MI After Myocardial Infarction**

1993

1994 This example demonstrates the following:

1995

- When it may not be possible to determine the NI margin because of the limitations of the data available.

1998

1999 By 1993, the effect of aspirin in preventing death after myocardial infarction had been
 2000 studied in six large randomized placebo-controlled clinical trials. A seventh trial, ISIS-2,
 2001 gave the drug during the first day after the AMI and is not included because it addressed a
 2002 different question. The results are summarized and presented in chronological order in Table
 2003 6.

2004

2005 **Table 6. Results of six placebo-controlled randomized studies (listed in chronological order) of the effect**
 2006 **of aspirin in preventing death after myocardial infarction**

Study	Year published	Aspirin		Placebo		Relative Risk (95% CI)
		N	Death rate	N	Death rate	
MRC-1	1974	615	8.0%	624	10.7%	0.74 (0.52, 1.05)
CDP	1976	758	5.8%	771	8.3%	0.70 (0.48, 1.01)
MRC-2	1979	832	12.2%	850	14.8%	0.83 (0.65, 1.05)
GASP	1978	317	10.1%	309	12.3%	0.82 (0.53, 1.28)
PARIS	1980	810	10.5%	406	12.8%	0.82 (0.59, 1.13)
AMIS	1980	2267	10.9%	2257	9.7%	1.12 (0.94, 1.33)

2007

2008 The results suggest:

2009

- (1) The effect of aspirin on mortality as measured by the relative risk seems to attenuate over the time the studies were conducted.
- (2) The largest trial, AMIS, showed a numerically adverse effect of aspirin.

2013

2014 The relative risk in the AMIS study is significantly different from the mean relative risk in
 2015 the remaining studies ($p \leq 0.005$). The validity of pooling the results of AMIS with those of
 2016 the remaining studies is therefore a concern. It would be invalid to exclude AMIS from the
 2017 meta-analyses because its effect differed from the effect in the remaining studies, unless there
 2018 were adequate clinical or scientific reasons for such exclusion. At a minimum, any meta-
 2019 analysis of all studies would need to reflect this heterogeneity by using a random-effect
 2020 analysis.

2021

2022 Although a fixed effect analysis of the six studies gives a point estimate of 0.91 (95% CI 0.82
 2023 to 1.02), the random-effects analysis gives a point estimate of 0.86 with 95% confidence
 2024 interval (0.69, 1.08). The effect of aspirin on prevention of death after myocardial infarction
 2025 in these historical studies is thus inconclusive (i.e., the upper bound of the 95% CI for effect
 2026 is > 1.0). Therefore, it would be difficult, indeed not really possible, to select aspirin as the

Contains Nonbinding Recommendations

Draft – Not for Implementation

2027 active control for evaluating the mortality effect of a test drug in a non-inferiority trial. Apart
2028 from this calculation, it seems difficult to accept an NI endpoint that is not supported by the
2029 largest of the six trials.

2030

2031 The same six studies can also be examined for the combined endpoint of death plus AMI in
2032 patients with recent AMI. This endpoint reflects the current physician-directed claim for
2033 aspirin based on the positive finding in two studies (MRC-2, PARIS).

2034

2035 **Table 7. Results of six placebo-controlled randomized studies of the effect of aspirin in secondary**
2036 **prevention of death or MI after myocardial infarction**

Study	Year published	Aspirin		Placebo		Relative Risk (95% CI)
		N	Event rate	N	Event rate	
MRC-1	1974	615	9.9%	624	13.1%	0.75 (0.55, 1.03)
CDP	1976	758	9.5%	771	12.5%	0.76 (0.57, 1.02)
MRC-2	1979	832	16.0%	850	22.2%	0.72 (0.59, 0.88)
GASP	1978	317	13.6%	309	17.5%	0.78 (0.54, 1.12)
PARIS	1980	810	17.4%	406	22.7%	0.77 (0.61, 0.97)
AMIS	1980	2267	18.6%	2257	19.2%	0.97 (0.86, 1.09)

2037

***the event rate of either group needs further verification from each article**

2038

2039 The results indicate that the effect of aspirin on death or MI after myocardial infarction is
2040 small to absent in the latest trial (AMIS). Random-effect analyses give, depending on the
2041 specific analysis, point estimates of the relative risk of 0.81-0.85, with 95% CI upper bounds
2042 of 0.96-1.02. The NI margin based on these six studies ranges from 4% to zero (without
2043 reducing it further to represent M_2) is so small that a trial to rule out loss at this effect would
2044 be unrealistically large. Again, as with the mortality endpoint, it would be troubling even to
2045 consider an NI approach when the largest and most recent trial showed no significant effect.

2046 **Example 4: Xeloda to Treat Metastatic Colorectal Cancer - the Synthesis Method**

2047

2048 This example of Xeloda for first-line treatment of metastatic colorectal cancer illustrates:

2049

2050 • The use of the synthesis method to demonstrate a loss of no more than 50% of the
2051 historical control treatment’s effect and a relaxation of this criterion when two NI studies
2052 are available.

2053

2054 • The use of supportive endpoints in the decision making process.

2055

2056 • The use of a conservative estimate of the control treatment effect size, because a subset
2057 of the available studies to estimate the margin was selected and the effect was measured
2058 relative to a previous standard of care instead of placebo.

2059

2060 The U.S. regulatory standard for first-line treatment of metastatic colorectal cancer, the use
2061 sought for Xeloda, is the demonstration of improvement in overall survival. Two separate
2062 clinical trials, each using an NI study design, compared Xeloda to a Mayo Clinic regimen of
2063 5-fluorouracil with leucovorin (5-FU+LV), the standard of care at the time. Xeloda is an oral
2064 fluoropyrimidine, while 5-fluorouracil (5-FU) is an infusional fluoropyrimidine

2065

2066 By itself, bolus 5-FU had not demonstrated a survival advantage in first-line metastatic
2067 colorectal cancer. But with the addition of leucovorin to bolus 5-FU, the combination had
2068 demonstrated improved survival. A systematic evaluation of approximately 30 studies that
2069 investigated the effect of adding leucovorin to a regimen of 5-FU identified ten clinical trials
2070 that compared a regimen of 5-FU+LV similar to the Mayo clinic regimen to 5-FU alone,
2071 thereby providing a measure of the effect of LV added to 5-FU, a conservative estimate of
2072 the overall effect of 5-FU+LV, as it is likely 5-FU has some effect.

2073

2074 Table 8 summarizes the overall survival results, using the metric “log hazard ratio” for the
2075 ten studies identified that addressed the comparison of interest.

2076

2077 **Table 8: Selected studies comparing 5FU to 5-FU+LV**

Study	Hazard Ratio ¹	Log Hazard Ratio ¹	Standard Error
Historical Study 1	1.35	.301	.232
Historical Study 2	1.26	.235	.188
Historical Study 3	0.78	-.253	.171
Historical Study 4	1.15	.143	.153
Historical Study 5	1.39	.329	.185
Historical Study 6	1.35	.300	.184
Historical Study 7	1.38	.324	.166
Historical Study 8	1.34	.294	.126
Historical Study 9	1.03	.0296	.165
Historical Study 10	1.95	.670	.172

2078

¹ All log hazard ratios are 5-FU/5-FU+LV

Contains Nonbinding Recommendations

Draft – Not for Implementation

2079 A random effects model applied to the survival results of these ten studies yielded the
2080 historical estimate of the 5-FU versus 5-FU+LV survival comparison of log hazard ratio of
2081 1.264 with a 95% confidence interval of (1.09, 1.46) and a log hazard ratio of 0.234. The NI
2082 margin is therefore 1.09 for a fixed margin approach ruling out M_1 .
2083

2084 A summary of the survival results based on the intent-to-treat populations for each of the two
2085 Xeloda NI trials is presented in Table 9. Study 2 rules out M_1 using a fixed margin approach,
2086 but Study 1 does not.
2087

2088 **Table 9: Summary of the survival results**

Study	Hazard Ratio ¹	Log Hazard Ratio ¹	Standard Error	95% CI for the Hazard Ratio ¹
NI Study 1	1.00	-0.0036	0.0868	(0.84, 1.18)
NI Study 2	0.92	-0.0844	0.0867	(0.78, 1.09)

2089 ¹ Hazard ratios and log hazard ratios are Xeloda/5-FU+LV
2090

2091 The clinical choice of how much of the effect on survival of 5-FU+LV should be shown not
2092 to be lost by Xeloda was determined to be 50%. The synthesis approach was used to analyze
2093 whether the NI criteria of 50% loss was met. This synthesis approach to the non-inferiority
2094 test procedure for each study combines the results of each NI study with the results from the
2095 random effects meta-analysis into a normalized test statistic.
2096

2097 Based on this NI synthesis test procedure, NI Study 1 failed to demonstrate that Xeloda
2098 retained at least 50% of the historical effect of 5-FU+LV versus 5-FU on overall survival, but
2099 NI study 2 did demonstrate such an effect. It was then decided to determine what percent
2100 retention might be satisfied by the data in a statistically persuasive way. By adapting the
2101 synthesis test procedure for retention of an arbitrary percent of the 5-FU+LV historical effect,
2102 it was determined that NI Study 1 demonstrated that Xeloda lost no more than 90% of the
2103 historical effect of 5-FU+LV on overall survival and that NI Study 2 demonstrated no more
2104 than a 39% loss of the historical effect.
2105

2106 The evidence of effectiveness of Xeloda was supported by the observation that the tumor
2107 response rates were statistically significantly greater for the Xeloda arm and the fact that
2108 Xeloda and 5-FU were structurally and pharmacologically very similar.
2109

2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154

REFERENCES - EXAMPLES

Example 1(A)

The Boston Area Anticoagulation Trial for Atrial Fibrillation Investigators (1990). “The Effect of Low-Dose Warfarin on the Risk of Stroke in Patients with Nonrheumatic Atrial Fibrillation.” *New Engl J Med* 323, 1505-1511.

Connolly, S.J., Laupacis, A., Gent, M., Roberts, R.S., Cairns, J.A., Joyner, C. (1991). “Canadian Atrial Fibrillation Anticoagulation (CAFA) Study.” *J Am Coll Cardiol* 18, 349-355.

EAFT (European Atrial Fibrillation Trial) Study Group (1993). “Secondary Prevention in Non-Rheumatic Atrial Fibrillation After Transient Ischemic Attack or Minor Stroke.” *Lancet* 342, 1255-1262.

Ezekowitz, M.D., Bridgers, S.L., James, K.E., Carliner, N.H., et al. (1992). “Warfarin in the Prevention of Stroke Associated with Nonrheumatic Atrial Fibrillation.” *N Engl J Med* 327, 1406-1412.

Food and Drug Administration, Dockets home page. Available at: http://www.fda.gov/ohrms/dockets/AC/04/briefing/2004-4069B1_07_FDA-Backgrounder-C-R-stat%20Review.pdf.

Halperin, J.L., Executive Steering Committee, SPORTIF III and V Study Investigators (2003). “Ximelagatran Compared with Warfarin for Prevention of Thromboembolism in Patients with Nonvalvular Atrial Fibrillation: Rationale, Objectives, and Design of a Pair of Clinical Studies and Baseline Patient Characteristics (SPORTIF III and V).” *Am Heart J* 146, 431-8.

Jackson, K., Gersh, B.J., Stockbridge, N., Fleiming, T.R., Temple, R., Califf, R.M., Connolly, S.J., Wallentin, L., Granger, C.B. (2005). Participants in the Duke Clinical Research Institute/American Heart Journal Expert Meeting on Antithrombotic Drug Development for Atrial Fibrillation (2008). “Antithrombotic Drug Development for Atrial Fibrillation: Proceedings.” Washington, D.C., July 25-27, 2005. *American Heart Journal* 155, 829-839.

Petersen, P., Boysen, G., Godtfredsen, J., Andersen, E.D., Andersen, B. (1989). “Placebo-Controlled, Randomised Trial of Warfarin and Aspirin for Prevention of Thromboembolic Complications in Chronic Atrial Fibrillation.” *The Lancet* 338, 175-179.

Stroke Prevention in Atrial Fibrillation Investigators (1991). “Stroke Prevention in Atrial Fibrillation Study: Final Results.” *Circulation* 84, 527-539.

Contains Nonbinding Recommendations

Draft – Not for Implementation

2155

2156 **Example 1(B) Refer to "General Reference" Section for synthesis methods.**

2157

2158 **Example 2**

2159

2160 Allais, J.M., Preheim, L.C., Cuevas, T.A., Roccaforte, J.S., Mellencamp, M.A., Bittner, M.J.
2161 (1988). "Randomized, Double-Blind Comparison of Ciprofloxacin and Trimethoprim
2162 Sulfamethoxazole for Complicated Urinary Tract Infections." *Antimicrob Agents*
2163 *Chemother.* 32(9), 1327-30.

2164

2165 Christiaens, T.C., De Meyere, M., Verschraegen, G., et al (2002). "Randomised Controlled
2166 Trial of Nitrofurantoin Versus Placebo in the Treatment of Uncomplicated Urinary Tract
2167 Infection in Adult Women." *Br J Gen Pract.* 52(482), 729-34.

2168

2169 DerSimonian, R., Laird, N. (1986), "Meta-Analysis in Clinical Trials," *Controlled Clinical*
2170 *Trials.* 7, 177-188.

2171

2172 Dubi, J., Chappuis, P., Darioli, R. (1982). "Treatment of Urinary Infection with a Single
2173 Dose of Co-trimoxazole Compared with a Single Dose of Amoxicillin and a Placebo."
2174 *Schweiz Med Wochenschr.* 12(3), 90-92.

2175

2176 Fang, G.D., Brennen, C., Wagener, M. et al (1991). "Use of Ciprofloxacin Versus Use of
2177 Aminoglycosides for Therapy of Complicated Urinary Tract Infection: Prospective,
2178 Randomized Clinical and Pharmacokinetic Study." *Antimicrob Agents Chemother.* 35(9),
2179 1849-55.

2180

2181 Ferry, S.A., Holm, S.E., Stenlund, H., Lundholm, R., Monsen, T.J. (2004). "The Natural
2182 Course of Uncomplicated Lower Urinary Tract Infection in Women Illustrated by a
2183 Randomized Placebo-Controlled Study." *Scan J Infect Dis.* 36, 296-301.

2184

2185 Ferry, S.A., Holm, S.E., Stenlund, H., Lundholm, R., Monsen, T.J. (2007). "Clinical and
2186 Bacteriological Outcome of Different Doses and Duration of Pivmecillinam Compared with
2187 Placebo Therapy of Uncomplicated Lower Urinary Tract Infection in Women: The LUTIW
2188 Project." *Scan J of Primary Health Care.* 25(1), 49-57.

2189

2190 Klimberg, I.W., Cox, C.E. 2nd, Fowler, C.L., King, W., Kim, S.S., Callery-D'Amico, S.
2191 (1998). "A Controlled Trial of Levofloxacin and Lomefloxacin in the Treatment of
2192 Complicated Urinary Tract Infection." *Urology.* 51(4), 610-5.

2193

2194 Peng, M.Y. (1999). "Randomized, Double-Blind, Comparative Study of Levofloxacin and
2195 Ofloxacin in the Treatment of Complicated Urinary Tract Infections." *J Microbiol Immunol*
2196 *Infect.* 32(1), 33-9.

2197

Contains Nonbinding Recommendations

Draft – Not for Implementation

2198 Talan, D.A., Stamm, W.E., Hooton, T.M. et al (2000). “Comparison of Ciprofloxacin (7
2199 Days) and Trimethoprim-Sulfamethoxazole (14 Days) for Acute Uncomplicated
2200 Pyelonephritis in Women.” *JAMA*. 283(12), 1583-1590.

2201

2202 **Example 3**

2203

2204 Aspirin Myocardial Infarction Study Research Group (1980). “A Randomized Controlled
2205 Trial of Aspirin in Persons Recovered from Myocardial Infarction.” *JAMA* 243, 661-669.

2206

2207 Breddin, K., Loew, D., Lechner, K., Uberia, E.W. (1979). “Secondary Prevention of
2208 Myocardial Infarction. Comparison of Acetylsalicylic Acid, Phenprocoumon and Placebo. A
2209 Multicenter Two-Year Prospective Study.” *Thrombosis and Haemostasis* 41, 225-236.

2210

2211 Coronary Drug Project Group (1976). “Aspirin in Coronary Heart Disease.” *Journal of*
2212 *Chronic Disease* 29, 625-642.

2213

2214 Elwood, P.C., Cochrane, A.L., Burr, M.L., Sweetnam, P.M., Williams, G., Welsby, E.,
2215 Hughes, S.J., Renton, R. (1974). “A Randomized Controlled Trial of Acetyl Salicylic Acid
2216 in the Secondary Prevention of Mortality from Myocardial Infarction.” *British Medical*
2217 *Journal* 1, 436-440.

2218

2219 Elwood, P.C., Sweetnam, P.M. (1979). “Aspirin and Secondary Mortality After Myocardial
2220 Infarction.” *Lancet* ii, 1313-1215.

2221

2222 Fleiss, J.L. (1993). “The Statistical Basis of Meta-Analysis.” *Statistical Methods in Medical*
2223 *Research* 2,121-145.

2224

2225 ISIS-2 Collaborative Group (1988). “Randomized Trial of Intravenous Streptokinase, Oral
2226 Aspirin, Both, or Neither Among 17187 Cases of Suspected Acute Myocardial Infarction:
2227 ISIS-2.” *Lancet* 2, 349-360.

2228

2229 Persantine-Aspirin Reinfarction Study Research Group (1980). “Persantine and Aspirin in
2230 Coronary Heart Disease.” *Circulation* 62, 449-461.

2231

2232 **Example 4**

2233

2234 FDA Guidance for Industry: Oncologic Drugs Advisory Committee Discussion on FDA
2235 Requirements for the Approval of New Drugs for Treatment of Colon and Rectal Cancer.

2236

2237 FDA Medical-Statistical review for Xeloda (NDA 20-896) dated April 23, 2001.

2238 (http://www.fda.gov/cder/foi/nda/2001/20896s6_Xeloda_Medr_Statr_P1.pdf).

2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283

GENERAL REFERENCES

Blackwelder, W.C. (1982). “Proving the Null Hypothesis in Clinical Trials.” *Controlled Clinical Trials* 3, 345-353.

Blackwelder, W.C. (2002). “Showing a Treatment is Good Because it is Not Bad: When Does “Noninferiority” Imply Effectiveness?” *Control Clinical Trials* 23, 52–54.

Brittain, E., Lin, D. (2005). “A Comparison of Intent-to-Treat and Per Protocol Results in Antibiotic Non-Inferiority Trials.” *Statistics in Medicine* 24, 1-10.

Brown, D., Day, S. (2007). Reply. *Statistics in Medicine* 26, 234-236.

CBER/FDA Memorandum (1999). Summary of CBER Considerations on Selected Aspects of Active Controlled Trial Design and Analysis for the Evaluation of Thrombolytics in Acute MI, June 1999.

Committee for Proprietary Medicinal Products (CPMP) (2000). Points to Consider on Switching Between Superiority and Non-Inferiority. <http://www.emea.europa.eu/pdfs/human/ewp/048299en.pdf>.

Committee for Medicinal Products for Human Use (CHMP) (2006). “Guideline on the Choice of the Non-Inferiority Margin.” *Statistics in Medicine* 25, 1628–1638.

Chow, S.C., Shao, J. (2006). “On Non-Inferiority Margin and Statistical Tests in Active Control Trial.” *Statistics in Medicine* 25, 1101–1113.

D’Agostino, R.B., Massaro, J.M., Sullivan, L. (2003). “Non-Inferiority Trials: Design Concepts and Issues – the Encounters of Academic Consultants in Statistics.” *Statistics in Medicine* 22, 169–186.

D’Agostino, R.B., Campbell, M., Greenhouse, J. (2005). “Non-Inferiority Trials: Continued Advancements in Concepts and Methodology.” *Statistics in Medicine* 25, 1097-1099.

DerSimonian, R., Laird, N. (1986). “Meta-Analysis in Clinical Trials.” *Controlled Clinical Trials* 7, 177-188.

Ellenberg, S.S., Temple, R. (2000). “Placebo-Controlled Trials and Active-Control Trials in the Evaluation of New Treatments - Part 2: Practical Issues and Specific Cases.” *Annals of Internal Medicine* 133, 464-470.

Fisher, L.D., Gent, M., Büller, H.R. (2001). “Active-Control Trials: How Would a New Agent Compare with Placebo? A Method Illustrated with Clopidogrel, Aspirin, and Placebo.” *American Heart Journal* 141: 26-32.

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 2284 Fleming, T.R. (1987). “Treatment Evaluation in Active Control Studies.” *Cancer Treatment*
2285 *Reports* 71, 1061-1064.
- 2286
- 2287 Fleming, T.R. (2000). “Design and Interpretation of Equivalence Trials.” *American Heart*
2288 *Journal* 139, S171-S176.
- 2289
- 2290 Follmann, D.A., Proschan, M.A. (1999). “Validity Inference in Random-Effects Meta-
2291 Analysis.” *Biometrics* 55, 732-737.
- 2292
- 2293 Freemantle, J., Cleland, J. Young, P., Mason, J., Harrison, J. (1999). “B Blockade After
2294 Myocardial Infarction: Systematic Review and Meta Regression Analysis.” *British Medical*
2295 *Journal* 318, 1730-1737.
- 2296
- 2297 Gould, A.L. (1991). “Another View of Active-Controlled Trials.” *Controlled Clinical Trials*
2298 12, 474-485.
- 2299
- 2300 Holmgren, E.B. (1999). “Establishing Equivalence by Showing That a Prespecified
2301 Percentage of the Effect of the Active Control Over Placebo is Maintained.” *Journal of*
2302 *Biopharmaceutical Statistics* 9(4), 651-659.
- 2303
- 2304 Hasselblad, V., Kong, D.F. (2001). “Statistical Methods for Comparison to Placebo in
2305 Active-Control Trials.” *Drug Information Journal* 35, 435-449.
- 2306
- 2307 Hauschke, D. (2001). “Choice of Delta: A Special Case.” *Drug Information Journal* 35,
2308 875-879.
- 2309
- 2310 Hauschke, D., Hothorn, L.A. (2007). Letter to the Editor: An Introductory Note to the
2311 CHMP Guidelines: Choice of the Non-Inferiority Margin and Data Monitoring Committees.
2312 *Statistics in Medicine* 26, 230-233.
- 2313
- 2314 Hauschke, D., Pigeot, I. (2005). “Establishing Efficacy of a New Experimental Treatment in
2315 the ‘Gold Standard’ Design (with discussions).” *Biometrical Journal* 47, 782-798.
- 2316
- 2317 Holmgren, E.B. (1999). “Establishing Equivalence by Showing that a Prespecified
2318 Percentage of the Effect of the Active Control Over Placebo is Maintained.” *Journal of*
2319 *Biopharmaceutical Statistics* 9, 651-659.
- 2320
- 2321 Hung, H.M.J., Wang, S.J., Tsong, Y., Lawrence, J., O’Neill, R.T. (2003). “Some
2322 Fundamental Issues with Non-Inferiority Testing in Active Controlled Clinical Trials.”
2323 *Statistics in Medicine* 22, 213-225.
- 2324
- 2325 Hung, H.M.J., Wang, S.J., O’Neill, R.T. (2005). “A Regulatory Perspective on Choice of
2326 Margin and Statistical Inference Issue in Non-Inferiority Trials.” *Biometrical Journal* 47,
2327 28-36.
- 2328

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 2329 Hung, H.M.J., Wang, S.J., O'Neill, R.T. (2008). "Non-Inferiority Trial." *Wiley*
2330 *Encyclopedia of Clinical Trials*. Wiley, New York.
2331
2332 Hung, H.M.J., Wang, S.J., O'Neill, R.T. (2007). "Issues with Statistical Risks for Testing
2333 Methods in Noninferiority Trial Without a Placebo Arm." *Journal of Biopharmaceutical*
2334 *Statistics* 17, 201-213.
2335
2336 International Conference on Harmonization: *Statistical Principles for Clinical Trials* (ICH
2337 E-9), Food and Drug Administration, DHHS, 1998.
2338
2339 International Conference on Harmonization: *Choice of Control Group and Related Design*
2340 *and Conduct Issues in Clinical Trials* (ICH E-10), Food and Drug Administration, DHHS,
2341 July 2000.
2342
2343 Jones, B., Jarvis, P., Lewis, J.A., Ebbutt AF (1996). "Trials to Assess Equivalence: the
2344 Importance of Rigorous Methods." *British Medical Journal* 313, 36-39.
2345
2346 Julious, S.A., Wang, S.J. (2008). "How Biased are Indirect Comparisons Particularly When
2347 Comparisons Are Made Over Time in Controlled Trials?" *Drug Information Journal* 42,
2348 625-633.
2349
2350 Koch, A., Röhmel, J. (2004). "Hypothesis Testing in the Gold Standard Design for Proving
2351 the Efficacy of an Experimental Treatment Relative to Placebo and a Reference." *Journal of*
2352 *Biopharmaceutical Statistics* 14, 315-325.
2353
2354 Kaul, S., Diamond, G.A. (2006). "Good Enough: A Primer on the Analysis and
2355 Interpretation of Non-Inferiority Trials." *Annals of Internal Medicine* 145, 62-69.
2356
2357 Lange, S., Freitag, G. (2005). "Choice of Delta: Requirements and Reality – Results of a
2358 Systematic Review." *Biometrical Journal* 47; 12-27.
2359
2360 Laster, L.L., Johnson, M.F., Kotler, M.L. (2006). "Non-Inferiority Trials: the 'at least as
2361 good as' Criterion with Dichotomous Data." *Statistics in Medicine* 25, 1115-1130.
2362
2363 Lawrence, J. (2005). "Some Remarks About the Analysis of Active Control Studies." *Biometrical Journal* 47, 616-622.
2364
2365
2366 Ng, T.H. (1993). "A Specification of Treatment Difference in the Design of Clinical Trials
2367 with Active Controls." *Drug Information Journal* 27, 705-719.
2368
2369 Ng, T.H. (2001). "Choice of Delta in Equivalence Testing." *Drug Information Journal* 35,
2370 1517-1527.
2371
2372 Ng, T.H. (2008). "Noninferiority Hypotheses and Choice of Noninferiority Margin." *Statistics in Medicine* 27, 5392-5406.
2373

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 2374
2375 Pledger, G., Hall, D.B. (1990). “Active Control Equivalence Studies: Do They Address the
2376 Efficacy Issue?” *Statistical Issues in Drug Research and Development*, Marcel Dekker, New
2377 York, 226-238.
2378
2379 Röhmel, J. (1998). “Therapeutic Equivalence Investigations: Statistical Considerations.”
2380 *Statistics in Medicine* 17, 1703-1714.
2381
2382 Rothmann, M., Li, N., Chen, G., Chi, G.Y.H., Temple, R.T., Tsou, H.H. (2003). “Non-
2383 Inferiority Methods for Mortality Trials.” *Statistics in Medicine* 22, 239-264.
2384
2385 Rothmann, M. (2005). “Type I Error Probabilities Based on Design-Stage Strategies with
2386 Applications to Noninferiority Trials.” *J. of Biopharmaceutical Statistics* 15; 109-127.
2387
2388 Sanchez, M.M., Chen, X. (2006). “Choosing the Analysis Population in Non-Inferiority
2389 Studies: Per Protocol or Intent-to-Treat.” *Statistics in Medicine* 25, 1169-1181.
2390
2391 Sheng, D., Kim, M.Y. (2006). “The Effects of Non-Compliance on Intent-to-Treat Analysis
2392 of Equivalence Trials.” *Statistics in Medicine* 25, 1183-1190.
2393
2394 Siegel, J.P. (2000). “Equivalence and Noninferiority Trials.” *American Heart Journal* 139:
2395 S166-S170.
2396
2397 Simon, R. (1999). “Bayesian Design and Analysis of Active Control Clinical Trials.”
2398 *Biometrics* 55, 484-487.
2399
2400 Snapinn, S.M. (2004). “Alternatives for Discounting in the Analysis of Noninferiority
2401 Trials.” *Journal of Biopharmaceutical Statistics* 14, 263-273.
2402
2403 Snapinn, S.M., Jiang, Q. (2008). “Controlling the Type I Error Rate in Non-Inferiority
2404 Trials.” *Statistics in Medicine* 27, 371–381.
2405
2406 Temple, R. (1987). “Difficulties in Evaluating Positive Control Trials.” *Proceedings of the*
2407 *Biopharmaceutical Section of American Statistical Association*, 1-7.
2408
2409 Temple R. (1996). “Problems in Interpreting Active Control Equivalence Trials.”
2410 *Accountability in Research* 4: 267-275.
2411
2412 Temple, R., Ellenberg, S.S. (2000). “Placebo-Controlled Trials and Active-Control Trials in
2413 the Evaluation of New Treatments - Part 1: Ethical and Scientific Issues.” *Annals of Internal*
2414 *Medicine* 133, 455-463.
2415
2416 Wang, S.J., Hung, H.M.J., Tsong, Y. (2002). “Utility and Pitfalls of Some Statistical
2417 Methods in Active Controlled Clinical Trials.” *Controlled Clinical Trials* 23, 15-28.
2418

Contains Nonbinding Recommendations

Draft – Not for Implementation

- 2419 Wang, S.J., Hung, H.M.J. (2003). “Assessment of Treatment Efficacy in Non-Inferiority
2420 Trials.” *Controlled Clinical Trials* 24, 147-155.
2421
2422 Wang S.J., Hung H.M.J. (2003). “TACT Method for Non-Inferiority Testing in Active
2423 Controlled Trials.” *Statistics in Medicine* 22; 227-238.
2424
2425 Wang, S.J., Hung, H.M.J., Tsong, Y. (2003). “Non-Inferiority Analysis in Active Controlled
2426 Clinical Trials.” *Encyclopedia of Biopharmaceutical Statistics, 2nd Edition*. Marcel Dekker,
2427 New York.
2428
2429 Wiens, B. (2002). “Choosing an Equivalence Limit for Non-Inferiority or Equivalence
2430 Studies.” *Controlled Clinical Trials* 23, 2-14.
2431
2432 Wiens, B. (2006). “Randomization as a Basis for Inference in Noninferiority Trials.”
2433 *Pharmaceutical Statistics* 5, 265-271.