Comments on ICH E17

General principles for planning and design of multi-regional clinical trials

Current Step 2 version, dated 6 May 2016

Comments by

the “German Region of the International Biometric Society (IBS-DR)” and

the “German Society for Medical Informatics, Biometry and Epidemiology (GMDS)”

30 January 2017

General comments:

1) The purpose of the guideline to promote MRCTs with the aim to reduce the number of separate clinical trials in different regions, and to avoid unnecessary duplication of studies is highly appreciated. This seems to suggest the application with only one well-planned and well-conducted MRCT allowing the investigation of consistency of results within this trial. Other guidance documents generally give the recommendation of two confirmatory clinical trials and formulate specific requirements in case of only one pivotal study. ICH E17 should give an explicit statement on the required number of MRCTs in different situations.

2) The guideline states that the determination of the overall sample size should ensure sufficient power for a test of the primary hypothesis in the total study population combining the data from all regions. The sample size allocation to regions should be determined such that clinically meaningful differences in treatment effects among regions can be described without substantially increasing the sample size requirements based on the primary hypothesis. The EMA Draft 'Guideline on the investigation of subgroups in confirmatory clinical trials’ supports an increase in sample size to investigate treatment effects by region. A harmonization of both guidance documents in this respect would be appreciated, and a more specific guidance on the required extent of consistency assessment in regions or pooled regions with limited sample size would be helpful.

3) If agreement on the primary endpoint cannot be reached among regulatory authorities, an MRCT is still recommended with a single protocol with endpoint-related sub-sections. This will lead to regulatory approvals based on different primary endpoints and no multiplicity adjustment is necessary. It has to be considered that many aspects of study planning as sample size planning and expectations about heterogeneity and effect modifiers largely depend on the primary endpoint and therefore more specific guidance for study planning and evaluation in this situation should be given.

Specific comments:

1) Chapter 1.4, line 69ff
   "For purposes of sample size planning and evaluation of consistency of treatment effects across geographic regions, some regions may be pooled at the design stage, if subjects in those regions are thought to be similar enough with respect to intrinsic and/or extrinsic factors relevant to the disease area and/or drug under study."
   Although we in principle agree with the guidance given, we feel that it would be worthwhile distinguishing more clearly between geographic regions and regions as used in design and analysis. Therefore, we propose to rephrase as follows: *"some geographic regions may be pooled at the design stage to regions as used in design and analysis"*.

2) Chapter 1.4, line 107ff
   "MRCTs may also serve as the basis for approval in regions not studied at the confirmatory stage through the extrapolation of study results."
   More specific guidance should be given on how study results could be extrapolated to regions not studied.

3) Chapter 2.1.1, Figure 1
   The question is, if the dashed submission line in the global strategy is realistic, considering the higher complexity in setting up MRCTs as compared to regional trials. It must be assumed that this will be shifted a bit to the right.

4) Chapter 2.2.4, line 354ff
   "It is in the interest of the sponsor to describe the specific advantages of the investigational product in terms of secondary endpoints as precisely as possible during the planning stage of MRCTs, to reduce the need for (and impact of) multiplicity adjustments for multiple endpoints, thereby improving the chance for successfully demonstrating the intended effect. Control of the Type I error across both primary and secondary endpoints may be required by some regulatory authorities."
   If control of the type I error for primary and secondary endpoints is required by some regulatory authorities, this has an impact on the necessary sample size and on the analysis. More specific guidance for study planning and evaluation should be given for the situation that no agreement between different authorities on multiple testing across primary and secondary endpoints can be reached.

5) Chapter 2.2.5, line 393ff
   "As stated in E9, the overall sample size is usually determined by the primary objective of the trial, stated in terms of study endpoints and specific hypotheses, as well as the size of the treatment effect to be detected …"
   This formulation would be correct if the test of a shifted null hypothesis were planned. Otherwise, the wording *"the size of the treatment effect to be detected"* should be changed to *"the size of the treatment effect assumed for sample size planning"* or something similar.

6) Chapter 2.2.7, line 602
The usefulness of the Funnel plot for the evaluation of regional consistency is not quite clear.

7) Chapter 2.2.7, line 623ff
"If the sample size in a region is so small that the estimates of effect are unreliable, the use of other methods should be considered, including the search for options to pool regions based on commonalities, or borrowing information from other regions or pooled regions using an appropriate statistical model."
A valid borrowing of information from a large region for a small region may be difficult. More specific guidance on appropriate statistical models and a critical discussion of the requirements would be helpful.

8) Chapter 2.2.7, line 609ff
"If subgroup differences (e.g., by gender) in treatment effects are observed, then an examination of whether the subgroup differences are consistent across regions or pooled regions is recommended."
This seems to be a good opportunity of MRCTs for the evaluation of consistency but it has to be considered that valid conclusions from the evaluation of these second order interactions may be difficult due to potentially small sample sizes.