

GUIDELINE ON CLINICAL TRIALS IN SMALL POPULATIONS

(CHMP/EWP/83561/2005, Draft, 17 March 2005)

German Region of the International Biometric Society

General Comments:

1. Definitions of what is meant by “small populations”, “very small populations”, “rare diseases” are neither provided in the introduction nor elsewhere in the paper.
2. The only clear guidance is to use more sophisticated rather than simpler methods for the analysis of clinical trials in smaller populations compared to larger populations. Otherwise, the guideline is a list of methods, which were or could be used with mentioning advantages and disadvantages, but without giving concrete advice what can be and what cannot be used when aiming to get a license. The same applies for endpoints and assumptions, regarding their validity.
3. Many of the statistical principles mentioned are not specific to clinical studies in small populations only, and should be done for studies in general.
4. The examples in the appendix are in most cases unclear and not well explained. It would be helpful to discuss different design and analysis approaches to the different examples in order to recommend which approaches should be avoided and which are preferred.

Specific Comments:

1. Page 3, 1. Introduction, 6th paragraph, last sentence

“This present guideline is exclusively intended for situations where established guidelines cannot be followed.”

Contrary to this statement Summary and Conclusions: bullet 2 (Page11) says: “Guidelines (ICH, CHMP and others) relating to common diseases are also applicable to rare diseases.”

What is correct?

2. Page 4, 2. Levels of Evidence, 3rd paragraph

“In very rare diseases, the combination of single case studies may be the only way to accumulate evidence. In such situations, treatment regimens and data collection may still be carried out in a controlled manner and this will add weight to the evidence.”



What is meant by “controlled” here? Is it equal to “prospective”? Or does it mean that a meta-analysis plan should be pre-specified and adhered to? What are “case studies” with respect to the “levels of evidence” list in the 1st paragraph: case reports, observational studies or anecdotal case-reports, or any of them? Please specify.

3. Page 6, 4. Choice of Endpoints, 3rd paragraph

“Relief of symptoms is also a useful clinical endpoint ... but it may not reflect slowing true disease progression or delaying death. Even in the absence of demonstration of benefit on a clinical endpoint, relief of symptom and the resulting patient reference may be a valuable study endpoint. However, this must always be on a disease and treatment-dependent justification.”

The guidance of this section whether “relief of symptoms” is a suitable endpoint remains unclear. Furthermore, this discussion is not different from studies in larger populations.

4. Page 6, 4. Choice of Endpoints, 4th paragraph

“If quality of life is measured, it should always be assessed using scales validated for the particular indication.”

It remains unclear how a validation of a scale could be achieved in small populations according to standard rules for quality of life instruments.

5. Page 6, 4. Choice of Endpoints, 6th paragraph

“Prediction in itself may not, however, be sufficient to attain the status of surrogate and a surrogate marker may not be sufficient to establish efficacy.

Further considerations should include:

- How closely changes in surrogate endpoint are linked to causing changes in a clinical or symptomatic endpoint.
- How much risk is associated with the therapy.
- What other therapies (if any) are available for the same condition.”

Why and in which way are the second and third points relevant for the question of a good surrogate endpoint?

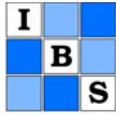
6. Page 7, 5. Choice of Control Groups, 3rd paragraph, and Page 8, 6.1. Design Stage, 4th paragraph

In these paragraphs dynamic/covariate-adaptive randomisation is recommended. In the CHMP Points to consider on adjustment of covariates, it is strongly recommended that these methods should be avoided. Please clarify this contradiction in CHMP Guidelines.

7. Page 7, 5. Choice of Control Groups, 4th paragraph

“Historical controls (...) might, under exceptional circumstances, be acceptable.”

No such circumstances are mentioned, but guidance should be given.



8. Page 7, 5. Choice of Control Groups, 4th paragraph

“If only active controlled studies are possible, then showing equivalence or non-inferiority may be difficult because assay sensitivity of the study cannot be assured and so obtaining a licence in these circumstances becomes extremely difficult.”

Does this mean it will be extremely difficult to get a licence in a very small patient population where a standard treatment is already available, the assay sensitivity regarding the standard has not been shown, yet, and the new drug does not claim to be better than the old one but to be non-inferior? Please specify.

9. Page 8, 6.1. Design Stage, 3rd paragraph

“Unreliability of one particular outcome can also be avoided by choosing another outcome (...), training of outcome assessors, and using multiple ratings.”:

This is not specific to clinical studies in small populations only and should be done for studies in general.

10. Page 8, 6.1. Design Stage, 8th paragraph, Last sentence

“The stratification factors must, however, be properly accounted for in the analysis.”

It is not clear which model should be chosen to account properly for stratification factors in case of adaptive methods.

11. Page 9, 6.1. Design Stage, 12th paragraph

“n-of-1 trials”

Is there any standard (published) statistical methodology for this type of studies? Wouldn't standard cross-over trials be a more prominent and methodologically better choice to make use of lower intra-individual variability compared to n-of-1-trials and shouldn't they therefore be explicitly mentioned in the guideline?

12. Page 10, 6.2. Data Analysis, 2nd paragraph

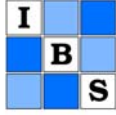
“Non-parametric methods”

What is meant with non-parametric methods here? It seems that only Bootstrap and Jackknife methods are meant whereas the more standard methods of non-parametric tests and estimation as well as of randomization/permutation tests are not mentioned explicitly. Guidance should be given which are the relevant ones for small populations.

13. Page 10, 6.2. Data Analysis, 2nd paragraph

“Non-parametric methods”

In this chapter, bootstrap methods are mentioned as an example for a statistical method that requires no assumption about the data distribution. In this context it should be mentioned that the same is true for permutation methods. It should however be made clear that the



application of such methods in principle does not allow to generalize the results from the investigated samples to the general population.

14. Page 10, 6.2. Data Analysis, 4th paragraph

"In cases where we may only get one study, it may be that careful modeling to determine which covariates and what functional form they take (e.g. linear, multiplicative, etc) is necessary."

This means that data-dependent analysis is permitted. This is in contradiction to CHMP Points to consider on adjustment of covariates. Are there any restrictions/guidelines on how to find the correct model? What must be pre-specified? It would be rather arduous to specify in advance every step and every procedure.

15. Page 11, 6.3. Interpretation of the Evidence

The reference Bradford-Hill is not mentioned in the reference section. Furthermore, the term 'controlled' seems to be used in the sense of 'non-interventional'. Otherwise it is unclear why criteria for determining causality from observational studies are discussed. There is more recent literature with regard to this topic.

16. Page 11, 7. Summary and Conclusions, 1st bullet

"The need for statistical efficiency should be weighed against the need for clinically interpretable results."

The contradiction between statistical efficiency and clinically interpretable result, which is discussed here, is not clear.

17. Page 11, 7. Summary and Conclusions, 3rd bullet

"In situations where obtaining controlled evidence on the efficacy and safety of a new treatment is not possible, the regulatory assessment may accept different approaches if they ensure that the patients' interests are protected."

The term controlled evidence remains unclear in this context and should be clarified.

18. Page 11, 7. Summary and Conclusions, 7th bullet

These aspects concerning patient registers mentioned here in the summary have nowhere been stated in the main text. Do any large patient registers exist which has documented their data in a GCP compliant way? GCP applies to clinical studies whereas patient register are non-interventional only. Therefore, GCP in total might be not the appropriate guidance for inspection of patient register.

19. Page 11, 7. Summary and Conclusions, 8th bullet

"When planned statistical (analysis) methods fail to show treatment effects, alternative approaches should be sought out (and preferably anticipated in the study protocol). Ideally



several methods should be applied and interpretation is helped if the results of different methodological/statistical approaches are in agreement.”

“Primary analyses” have nowhere been mentioned in the guideline. So for small population studies has a primary analysis to be pre-defined and followed, or would there be a variety of different pre-defined analyses which may be differently important but where none of the analyses would have the role of the primary analysis (“success or no success”) in larger population studies?

20. Page 13, Appendix, 7th and 8th paragraph

“Response adaptive methods”

Example of heavily criticized trial (1:11 subjects) and other example where study starts with randomizing 1:1:1: There is no clear statement in the guideline on the second example and even not on the first one whether a license was, can be or would be obtained by such studies.

21. Page 15, Appendix, 15th paragraph

“Non-parametric methods”

The example remains unclear.

22. Page 15, Appendix, 17th paragraph

Bayesian methods are recommended as a tool for the analysis of trial data from rare diseases. Reference is made to an article from Tan et al (Strategy for randomised clinical trials in rare cancers: BMJ 2003, 327;47-49). The paper by Tan (2003) serves as a background document. Unfortunately, however, the authors of the draft guideline misunderstood the paper when stating: “they (Tan et al) describe how to use predefined scenarios within the frame of a sensitivity analysis: they suggest a sceptical prior distribution (it is assumed that the new therapy is even worse than standard treatment); a neutral prior distribution (the new therapy has no effect at all); an enthusiastic prior distribution (the new treatment has a predefined realistic effect).”

Tan et al. do not talk about prior distributions but on prior and posterior distributions and likelihood "data" for enthusiastic, neutral, and sceptical scenarios. The prior distribution in this example assumes that the two treatments are identical (log hazard ratio=0).

It would come close to a disaster for the acceptability of Bayesian methods in clinical trials if indeed the presented view on what constitutes a sceptical, neutral or enthusiastic prior found its way into regulatory guidance.

To recall, Spiegelhalter et al. (Bayesian Approaches to Clinical Trials and Health Care Evaluation. Wiley 2004) explain the sceptical prior distribution by stating: “In a study where the aim is to demonstrate superiority the sceptical prior distribution has mean zero and the precision is such that the prior probability is low (e.g. 5%) that the true benefit exceeds the expected mean under the alternative.”