26 February 2015
EMA/CHMP/295050/2013
Committee for Medicinal Products for Human Use (CHMP)

Guideline on adjustment for baseline covariates in clinical trials

Draft Agreed by Biostatistics Working Party (BSWP)	March 2013
Adoption by CHMP for release for consultation	30 May 2013
Start of public consultation	28 June 2013
End of consultation (deadline for comments)	31 December 2013
Agreed by BSWP	January 2015
Adopted by CHMP	26 February 2015
Date for coming into effect	01 September 2015

This guideline replaces Points to Consider on adjustment for baseline covariates (CPMP/EWP/2863/99)

Comments should be provided using this template. The completed comments form should be sent to Biostatistics@ema.europa.eu.

Keywords	*Baseline covariates, stratification, dynamic allocation, multicentre trials, baseline imbalance*

30 Churchill Place ● Canary Wharf ● London E14 5EU ● United Kingdom
Telephone +44 (0)20 3660 6000 **Facsimile** +44 (0)20 3660 5555
Send a question via our website www.ema.europa.eu/contact

Guideline on adjustment for baseline covariates in clinical trials

Table of contents

Executive summary

Baseline covariates impact the outcome in many clinical trials. Although baseline adjustment is not always necessary, in case of a strong or moderate association between a baseline covariate(s) and the primary outcome measure, adjustment for such covariate(s) generally improves the efficiency of the analysis and avoids conditional bias from chance covariate imbalance.

Baseline covariates may be accounted for at the design stage of a clinical trial and/or in the statistical analysis. When dealing with baseline covariates the following recommendations are made:

- Stratification may be used to ensure balance of treatments across covariates; it may also be used for administrative reasons (*e.g.* block in the case of block randomisation). The factors that are the basis of stratification should normally be included as covariates or stratification variables in the primary outcome model, except where stratification was done purely for an administrative reason.
- Variables known *a priori* to be strongly, or at least moderately, associated with the primary outcome and/or variables for which there is a strong clinical rationale for such an association should also be considered as covariates in the primary analysis. The variables selected on this basis should be pre-specified in the protocol.
- Baseline imbalance observed *post hoc* should not be considered an appropriate reason for including a variable as a covariate in the primary analysis. However, conducting exploratory analyses including such variables when large baseline imbalances are observed might be helpful to assess the robustness of the primary analysis.
- Variables measured after randomisation and so potentially affected by the treatment should not be included as covariates in the primary analysis.
- If a baseline value of a continuous primary outcome measure is available, then this should usually be included as a covariate. This applies whether the primary outcome variable is defined as the 'raw outcome' or as the 'change from baseline'.
- Covariates to be included in the primary analysis must be pre-specified in the protocol.
- Only a few covariates should be included in a primary analysis. Although larger data sets may support more covariates than smaller ones, justification for including each of the covariates should be provided.
- In the absence of prior knowledge, a simple functional form (usually either linearity or categorising a continuous scale) should be assumed for the relationship between a continuous covariate and the outcome variable.
- The validity of model assumptions must be checked when assessing the results. This is particularly important for generalised linear or non-linear models where mis-specification could lead to incorrect estimates of the treatment effect. Even under ordinary linear models, some attention should be paid to the possible influence of extreme outlying values.
- Whenever adjusted analyses are presented, results of the treatment effect in subgroups formed by the covariates (appropriately categorised, if relevant) should be presented to enable an assessment of the model assumptions.
- Sensitivity analyses should be pre-planned and presented to investigate the robustness of the primary analysis. Discrepancies should be discussed and explained. In the presence of important differences that cannot be logically explained – for example, between the results of adjusted and unadjusted analyses – the interpretation of the trial could be seriously affected.

- The primary model should not include treatment by covariate interactions. If substantial interactions are expected *a priori*, the trial should be designed to allow separate estimates of the treatment effects in specific subgroups.
- Exploratory analyses may be carried out to improve the understanding of covariates not included in the primary analysis, and to help the sponsor with the ongoing development of the drug.
- In case of missing values in baseline covariates the principles for dealing with missing values as outlined e.g. in the Guideline on missing data in confirmatory clinical trials (EMA/CPMP/EWP/1776/99 Rev. 1) applies.
- A primary analysis, unambiguously pre-specified in the protocol, correctly carried out and interpreted, should support the conclusions which are drawn from the trial. Since there may be a number of alternative valid analyses, results based on pre-specified analyses will carry most credibility.

Besides editorial changes the major change with this revision of the Guideline relates to the use of dynamic allocation methods.

1. Introduction

The note for guidance on statistical principles for clinical trials (ICH E9) briefly addresses the problem of adjustment for covariates. It advises experimenters 'to identify the covariates expected to have an important influence on the primary outcome' and to specify 'how to account for them in the analysis in order to improve precision and to compensate for any lack of balance between groups'. It also cautions against adjusting for 'covariates measured after randomisation because they may be affected by the treatments'.

A baseline covariate in the context of this guideline is defined as a qualitative factor or a quantitative variable measured or observed before randomisation and expected to influence the primary outcome variable to be analysed.

There are many types of baseline covariates and their nature depends upon the context of the study. They may be demographic variables such as age or weight, disease characteristics such as duration or severity, true prognostic factors for which there is a commonly accepted pathophysiological rationale, or factors such as centre or investigator. Quite commonly baseline values of the primary outcome variable are also available.

A baseline covariate can be considered at two stages in a clinical trial: it can be accounted for within the randomisation process (typically by using stratified randomisation) and/or it can be adjusted for in the analysis.

There are many different techniques for adjusting for baseline covariates, the choice of which often depends on the nature of the covariate and outcome variable. Methods commonly used are analysis of variance or analysis of covariance (when the primary outcome is quantitative), logistic regression (for binary or categorical data), and Cox regression (for time-to-event data) and/or stratified analyses (*e.g.* Cochran-Mantel-Haenszel test in case of binary data).

The guideline aims to clarify when and why baseline covariates should be included in the primary analysis that will be specified in the protocol, and how the results in the study report should be presented and interpreted. A question that is often encountered is whether the adjusted or unadjusted analysis should be declared as primary in the protocol. This guidance document addresses that critical issue.

2. Scope

This guideline is intended to provide advice on how to address important baseline covariates in designing, analysing and reporting clinical trials. Its content is mostly concerned with confirmatory randomised trials.

Non-randomised trials, such as observational studies, as well as technical and theoretical aspects of methods to account for covariates and discussions on the clinical relevance of particular choices of covariates are outside the scope of this guideline.

3. Legal basis and relevant guidelines

The Guideline should be read in conjunction with Annex I to Directive 2001/83/EC, as amended, and all other relevant EU and ICH guidelines. These include, but are not limited to:

- CPMP/ICH/363/96, ICH Topic E9 Step 4 Note for Guidance on Statistical Principles for Clinical Trials.
- Guideline on missing data in confirmatory clinical trials (EMA/CPMP/EWP/1776/99 Rev. 1)
- Points to consider on multiplicity issues in clinical trials (CPMP/EWP/908/99)

4. Design consideration

4.1. Stratification

Randomisation is expected to balance treatment groups among the covariate levels on average but, in practice, it is not unusual to observe imbalances *post hoc*. Such imbalances are of particular concern if they favour the experimental group. Stratified randomisation is often used to reduce the likelihood of such imbalances between treatment groups within the levels of specified covariates (generally qualitative covariates or categorised quantitative covariates).

Additional reasons why stratified designs are used include:

- Balance of treatment groups with respect to one or more specific prognostic covariates can enhance the credibility of the results of the trial.
- Stratification might improve the efficiency of the estimation of the treatment effect, especially for small or even moderately sized trials. Stratification at the stage of randomisation and adjustment for covariates in the analysis may be seen as complementary methods of accounting for covariates.
- If the effect of treatment is expected to vary substantially across important pre-specified subgroups (for example, age groups or race), then stratifying for these subgroups can help in interpreting the treatment effect and its consistency across these subgroups. This can also enhance the credibility of some subgroup analyses that are *a priori* of high interest. For further details refer to regulatory documents on subgroup analysis.
- Stratification may sometimes be used for reasons of administrative convenience.

Stratification can become overwhelming if there are many influential covariates or covariates with many strata in the trial. This is particularly true for small trials where stratification on more than a few covariates is often not feasible due to small sample sizes within strata. Even in large trials, although theoretically possible to stratify by many factors, the number of factors should be restricted to the most clinically important and/or strongly prognostic covariates. With an increasing number of strata

the chance of empty / infrequently occupied strata increases, thus the targeted treatment allocation within strata might not be achieved. Furthermore, a huge number of strata might impose problems with the analysis (see 6.2).

4.2. Dynamic Allocation

As stated above, stratification for more than a few prognostic factors is not always possible, especially for small trials. In this situation, techniques of dynamic allocation are sometimes used to achieve balance across several factors simultaneously. The (simplified) idea behind dynamic allocation is to measure the imbalance marginally over each prognostic factor and to minimize *e.g.* the (weighted) sum of imbalances over all prognostic factors, thus maximising overall balance. However, dynamic allocation does not guarantee balance within combinations of prognostic factors. Deterministic schemes should be avoided.

Possible implications of dynamic allocation methods on the analysis *e.g.* with regard to bias and Type I error control should be carefully considered, taking into account that for some situations (*e.g.* planned unbalanced treatment allocation that allows to change the allocation ratio at every allocation) it has been shown that in case of dynamic treatment allocation conventional statistical methods do not always control the Type I error. To properly account for such problems the use of re-randomisation methods in the analysis should be considered.

4.3. Multicentre trials

Multicentre trials are often stratified by centre either for practical reasons or because centre is expected to be confounded with other known or unknown prognostic factors. When multicentre trials are not stratified by centre, then the reason for doing so should be explained in the protocol.

If a multicentre trial is not stratified for centre (*e.g.* when the number of patients within many centres is expected to be very small), it should be considered whether randomisation could be stratified by, for example, country or region. Such a choice might be driven by similarities in co-medication, palliative care or other factors that might make stratification advisable. The reasons and justification for the choice should be described in the protocol.

5. Criteria for including OR excluding a covariate in the primary analysis

5.1. Association with the Primary Outcome

The main reason to include a covariate in the analysis of a trial is evidence of strong or moderate association between the covariate and the primary outcome measure. Adjustment for such covariates generally improves the efficiency of the analysis and hence produces stronger and more precise evidence (smaller *P*-values and narrower confidence intervals) of an effect. However, it should be emphasised that simply producing smaller *P*-values may not be sufficient to produce convincing evidence of a clinically useful effect: the size of the treatment effect and its consistency across levels of covariates will always be important considerations.

Known or expected associations with the primary outcome variable should be justified on the basis of previous evidence (*e.g.* data from previous trials) and/or on clinical grounds. The reasons for including a covariate in the primary analysis should be explicitly stated in the protocol.

5.2. Stratification

The primary analysis should reflect the restriction on the randomisation implied by the stratification. For this reason, stratification variables, if not solely used for administrative reasons, should usually be included as covariates or stratification variables in the primary analysis regardless of their prognostic value. Any mismatch of non-administrative covariates between stratification and adjustment in the primary analysis must be explained and justified. This includes the use of covariates not stratified for in the randomisation.

5.3. Multicentre trials

In multicentre trials randomisation might be stratified by centre, country and/or region. The stratification variables used for randomisation should be adjusted for in the primary analysis.

If centre was used for stratification in a multicentre trial problems might arise in case of many centres recruiting small numbers of patients ('small centres'). Adjusting for many small centres might be possible but raises analytical problems for which there is no best solution. Analyses either ignoring centres used in the randomisation or adjusting for a large number of small centres might lead to unreliable estimates of the treatment effect and *P*-values that may be either too large or too small. Furthermore, pooling small centres to form one centre of size comparable to that of other centres has little or no scientific justification. If an applicant chooses not to include centre in the analysis when it was included in the randomisation scheme, they should explain why and demonstrate through sensitivity analyses that the trial conclusions are not substantially affected because of this.

5.4. Baseline imbalance observed post hoc

A pronounced baseline imbalance is not expected *a priori* in a randomised trial: if the randomisation process has worked correctly, any observed imbalance is likely to be a random phenomenon. Therefore, if a baseline imbalance is observed this should not be considered an appropriate reason to include this baseline measure as a covariate in the primary analysis. In case the baseline imbalance is for a prognostic factor, sensitivity analyses including the baseline measure as a covariate should be performed in order to assess the robustness of the primary analysis.

5.5. Covariates affected by the treatment allocation

A covariate that may be affected by the allocated treatment (for example, a covariate measured after randomisation such as duration of treatment, level of compliance or use of rescue medication) should not be included in the primary analysis of a confirmatory trial. When a covariate is affected by the treatment either through direct causation or through association with another factor, the adjustment may hide or exaggerate the treatment effect. It therefore makes the treatment effect difficult to interpret. However, such covariates (*e.g.* duration of treatment) might be included in secondary (exploratory) analyses and might offer the sponsor useful insights during the drug development process. Alternatively, subgroup analyses might offer similar insights.

5.6. 'Change from baseline' analyses

When the primary analysis is based on a continuous outcome there is commonly the choice of whether to use the raw outcome variable or the change from baseline as the primary endpoint. Whichever of these endpoints is chosen, the baseline value should be included as a covariate in the primary analysis. The use of change from baseline with adjustment for baseline is generally more precise than change of baseline without adjustment. Note that when the baseline is included as a covariate in a standard

linear model, the estimated treatment effects are identical for both 'change from baseline' (on an additive scale) and the 'raw outcome' analysis. Consequently if the appropriate adjustment is done, then the choice of endpoint becomes solely an issue of interpretability.

6. Specification of the primary analysis

6.1. General considerations

Covariates to be included in the primary analysis *must* be pre-specified in the protocol. When a confirmatory (typically phase III) trial starts, the important covariates should have already been identified through previous trials and other available evidence. However, if the state of knowledge changes between the writing of the protocol and the completion of the study it may be appropriate to re-consider and update the description of the analysis in the statistical analysis plan prior to any unblinding. The justification (at this time) for including new covariates (or excluding others that were previously identified) should be stated unambiguously. Both clinical and statistical justifications should be considered. It might be of added value to present also the initially planned analysis. When there is a lack of established prior knowledge, it is safer to use a simple model with no, or only a few, covariates. In general, analyses including many covariates will always be less convincing than analyses with fewer, well-chosen, covariates.

The nature and the number of covariates included in the analysis may affect the interpretation of the analysis, especially in non-linear models. In such models the adjusted parameters and unadjusted parameters have different interpretations: it is essential that in any presentation of adjusted analyses, the applicant clearly and accurately explains the meaning of the estimated effect size.

Methods that retrospectively select covariates by choosing those that are most strongly associated with the primary outcome (often called 'variable selection methods') should be avoided in confirmatory clinical trials. The clinical and statistical relevance of a covariate should be assessed and justified from a source other than the current dataset.

In some cases, not all of the relevant sensitivity analyses for a particular study can be anticipated in the protocol. However some sensitivity analyses should be pre-planned to establish whether the conclusions drawn from the primary analysis are robust. In particular, sensitivity analyses should be designed to test specific assumptions about covariates.

6.2. Number of covariates in the analysis

No more than a few covariates should be included in the primary analysis. Even though methods of adjustment, such as analysis of covariance, can theoretically adjust for a large number of covariates it is safer to pre-specify a simple model. Results based on such a model are more likely to be numerically stable, the assumptions underpinning the statistical model are easier to validate and generalisability of the results may be improved.

There is no formal rule for specifying the maximum number of covariates that can be included in any analysis, although larger trials might tolerate more covariates than smaller trials. Potential covariates are often strongly correlated and so knowledge of the correlation can be a useful basis for eliminating some stratification variables at the planning stage. Clinical considerations should be taken into account when doing this.

Limitations should be placed on the number of covariates included in the statistical model and on the total number of parameters. Categorical covariates with many levels may lead to a loss of efficiency. For such covariates, strategies to combine categories or to carry out alternative sensitivity analyses

should be pre-specified in the protocol.

6.3. Relationship between covariates and the primary outcome

The aim of a randomised clinical trial is not to determine the true relationship between covariates and the primary outcome variable but to provide an unbiased estimate of the true difference between the treatments.

The true relationship between covariates and the primary outcome variable is often unknown (*e.g.* whether there is a linear or quadratic relationship between a covariate and the outcome). Thus the behaviour of the analysis model under mis-specification should be considered when defining the analysis model. In the absence of any well-established prior knowledge about the relationship between the covariates and the outcome (which is often the case in clinical trials) the model should use a simple form. For example, when the covariate is continuous, then the model might be based on a linear relationship between the covariate and outcome (on whichever additive scale to be used), or on a categorisation of the covariate into a few levels, the number of levels depending upon the sample size. In such a case, the rules for determining how the categories will be described should be pre-specified and sensitivity analyses conducted to ensure subsequent conclusions are not highly dependent on the categories selected.

If there is well-established prior information from previous studies about how the covariates are related to the outcome, then the primary model should incorporate this information. The functional form that relates the covariates to the outcome should be pre-specified and justified in the protocol. Nonparametric regression methods may be applied which avoid assumptions about the relationship between the dependent and independent variables. However, in these cases, it is important that appropriate estimates of the size of the treatment effect are still attainable, not just the calculation of significance levels.

In addition to the functional form relating covariates to the outcome, attention should be paid to outlying values of either the covariates or the outcome variable as these may have undue influence on the results. If the possibility of outlying values is foreseen, then their influence can be minimised by pre-planning suitable robust methods (*e.g.* non-parametric analyses).

6.4. Treatment by covariate interaction

This has already been addressed in ICH E9 and is not an issue specifically related to adjustment for covariates. The fact that the treatment effect may be different depending on the baseline value of a covariate is a matter for concern whether adjustment for this covariate is considered or not.

If there is no reason to suspect an interaction between treatment and a covariate then the primary analysis should only include the main effects for treatment and covariate. Conversely, if a substantial treatment by covariate interaction is suspected at the design stage, then stratified randomisation and/or subgroup analyses should be pre-planned accordingly. For details refer to further regulatory documents dealing with multiplicity and subgroup analysis respectively.

6.5. Missing covariate information

There might be situations where covariate information is lacking in individual patients. Ignoring these patients in the analysis violates the ITT principle and is not considered acceptable in general. Therefore, in case covariates are part of the analysis model for the primary outcome it should be considered at the planning stage of the trial how to deal with missing covariate information. The principles to deal with missing values as outlined in the corresponding guideline on missing values in

confirmatory clinical trials should be followed.

7. Report of the results

7.1. General considerations

If the key covariates were specified clearly in the protocol and the analysis was correctly performed and interpreted, then appropriate conclusions can be safely drawn. However, if the covariates and the method of adjustment for them were not specified unambiguously, then a number of alternative analyses may be equally valid. It will be difficult for the applicant to argue *post hoc* that a particular analysis is not data driven.

7.2. Baseline comparisons

Statistical testing for baseline imbalance has no role in a trial where the handling of the randomisation and blinding has been fully satisfactory. Baseline summaries with respect to the main covariates should be presented and discussed from a clinical point of view as any observed imbalance will be a random phenomenon.

If treatment allocation was not random then any resulting bias cannot be corrected by statistical adjustment. The appropriate actions (possibly excluding some patients or centres) will follow from investigations into the cause of the imbalance. The results should be interpreted very cautiously in such cases.

When there is some imbalance between the treatment groups in a baseline covariate that is solely due to chance then adjusted treatment effects may account for this observed imbalance when unadjusted analyses may not. If the imbalance is such that the experimental group has a better prognosis than the control group, then adjusting for the imbalance is particularly important. Sensitivity analyses should be provided to demonstrate that any observed positive treatment effect is not solely explained by imbalances at baseline in any of the covariates.

In the unlikely case of a very strong baseline imbalance, no adjustment may be sufficiently convincing to restore the reliability of the results. However, a strong baseline imbalance in a variable (not necessarily a pre-specified covariate) may also be a reason for including that variable as a covariate in a sensitivity analysis to allow assessment of the robustness of the conclusions drawn from the primary analysis.

7.3. Treatment by covariate interaction

The primary analysis should include only the covariates pre-specified in the protocol and no treatment by covariate interaction terms. However, treatment by covariate interactions should be explored, as recommended in the ICH E9 guideline. Tests for interactions often lack statistical power and the absence of statistical evidence of an interaction is not evidence that there is no clinically relevant interaction. Conversely, an interaction cannot be considered as relevant on the sole basis of a significant test for interaction. Assessment of interaction terms based on statistical significance tests is therefore of little value. Alternative approaches to assess possible treatment by covariate interactions (*e.g.* presenting results by stratum) should be pre-planned.

If some treatment by covariate interactions turn out to be large from a clinical point of view or significant from a statistical point of view, this provides evidence that the effect of treatment may vary across subgroups. These findings should be examined carefully; conclusions based on the primary analysis (with no interaction) should be interpreted cautiously and commented on. If the observed

interaction is particularly large in size or qualitative in nature, then interpretation of the overall results of the trial may become impossible.

7.4. Validity of the model assumptions

Mis-specification of the analysis model could lead to incorrect estimates of the treatment effect. Thus, assumptions must be checked carefully (*e.g.* by assessing residuals) and any findings indicating model mis-specification should be presented in the final study report. If the model assumptions do not hold, alternative analyses (ideally pre-specified in the protocol) should be proposed and justified on clear statistical and clinical grounds in order to allow for an assessment of the robustness of the primary conclusion.

7.5. Sensitivity analyses

Alternative analyses should always be presented to confirm that the conclusions of the study are not sensitive to the choice of covariates included or the choice of the relationship between covariates and outcome that has been assumed. Findings based on these sensitivity analyses should normally be considered exploratory but necessary to support the primary analysis.

For ordinary linear models, adjusted estimates of the treatment effect should be compared to unadjusted estimates. The estimates of the size of the treatment effect would be expected to be similar although not necessarily identical. Since there is generally an expected gain in efficiency with the adjusted analysis, a less significant result for an unadjusted analysis is not necessarily cause for concern. Conversely, if there are strong discrepancies between the conclusions drawn from adjusted and unadjusted analyses, these should be discussed and interpreted whenever possible. If the conclusions from the primary analysis and the sensitivity analyses are very different in terms of clinical and statistical significance and the difference cannot be explained by *e.g.* imbalances between treatment groups in the covariates, then the results of the trial could become inconclusive.

For generalised linear models or non-linear models, adjusted and unadjusted treatment effects may not have the same interpretation and, sometimes, different results may be obtained from adjusted and unadjusted analyses. Thus, the choice of the appropriate covariates and the pre-specification of the primary model are critically important.