EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

19 September 2012
EMA/868942/2011
Good Clinical Practice Inspectors Working Group (GCP IWG)

Points to consider on GCP inspection findings and the benefit-risk balance

Agreed by GCP IWG	12 September 2012
Adopted by CHMP	8 October 2012

7 Westferry Circus ● Canary Wharf ● London E14 4HB ● United Kingdom
Telephone +44 (0)20 7418 8400 **Facsimile** +44 (0)20 7418 8416
E-mail info@ema.europa.eu **Website** www.ema.europa.eu
An agency of the European Union

Points to consider on GCP inspection findings and the benefit-risk balance

Table of contents

General considerations

Good Clinical Practice is a set of internationally recognised ethical and scientific quality requirements which must be observed for designing, conducting, recording and reporting clinical trials that involve the participation of human subjects. Compliance with this good practice provides assurance that the rights, safety and well-being of trial subjects are protected, and that the results of the clinical trials are credible (EU Directive 2001/20/EC in Article 1 (Scope)).

GCP inspection findings in studies forming the basis for an application for marketing authorizations for new medicinal products or for extending indications to already authorized products have two main implications: Some findings point to GCP non-compliance issues affecting the safety, well-being and the rights of participating subjects; other findings might (also) question the quality and integrity of the data and thus the interpretability of the study results and consequently make it difficult to use the data of study in the evaluation of the benefit-risk balance.

GCP inspectors and clinical assessors have different roles in the overall regulatory process of evaluating new medicines. Hence, it should be acknowledged that the focus of the GCP inspectors and clinical assessors is different and as a consequence the evaluation of the significance of the findings may also differ. For the assessors, the focus is on the particular medicine under assessment to ensure that the benefit-risk balance is favourable before licensing and that the overall ethical conduct is acceptable. However for inspectors, the focus is not only on the individual medicine, but also on the protection of the rights, safety and wellbeing of the trial subjects and on aspects related to whether the sponsor has a robust quality system in place to guarantee the quality of the data of the trial(s) inspected and also of future upcoming applications (since it is not possible to inspect all of them).

The objective of this document is to assist inspectors and assessors in evaluating the consequences of inspection findings in relation to the benefit-risk balance. It should help inspectors in drafting the inspection reports and improve mutual understanding between inspectors and assessors in order to effectively aid clinical assessors, rapporteurs and ultimately the CHMP in their scientific evaluation of the benefit-risk balance. For this purpose, it is important to distinguish those findings that are likely to have an impact on the benefit-risk evaluation and those that are not. In this document, an attempt to rate inspection findings by their importance to the benefit-risk evaluation is made. Three categories are used:

- inspection findings which are likely to influence the benefit-risk evaluation;
- inspection findings which may influence the benefit-risk evaluation;
- inspection findings which are less likely to influence the benefit-risk evaluation.

This classification should be seen as a general principle which may be departed from when appropriate. It is intended to stimulate a thought process when evaluating inspection findings in terms of their impact on the benefit-risk assessment, but the terminology used for the three categories is not to be viewed as a formal grading system and do not need to be used, e.g. in the inspection reports or assessment reports. Obviously, findings that are considered to influence the benefit-risk evaluation will often be graded as “critical” as per the “official” grading of inspection findings (Procedure for reporting of GCP inspections requested by the EMEA, EMEA/INS/GCP/197226/2005). Critical inspection findings are conditions, practices or processes that adversely affect the rights, safety or well being of the subjects and/or the quality and integrity of data.

Each individual inspection finding should be interpreted on a case-by-case basis and viewed in the context of the clinical study, the development programme and the totality of the available information

about the medicinal product. The potential impact of the findings on the benefit-risk assessment should be analysed and discussed by the applicant.

As highlighted in the next section, major ethical flaws – even if not directly influencing the benefit-risk assessment – should have an impact on the final conclusions about approvability of an application.

This document applies to GCP inspections resulting from both routine and triggered inspections and has to be viewed in the perspective of other guidelines addressing GCP principles and GCP inspections.

It is the intention that this document will be updated based on further experience/evolving knowledge.

1. Ethics

The EU legislation requires not only valid clinical data for the scientific evaluation of the benefit-risk balance, but also ethical conduct of the clinical development programme in order to ensure that the rights, safety and well being of the trial subjects are protected. GCP inspection findings – even if not directly influencing the benefit-risk balance - will still be important if they raise serious questions about the rights, safety and well-being of trial subjects and hence the overall ethical conduct of the study. It is an obligation of clinical assessors, rapporteurs and the CHMP also to assess the ethics of a clinical development programme, and major ethical flaws should have an impact on the final conclusions about approvability of an application. Consequently, ethical misconduct could result in rejection of the application.

Some inspection findings mentioned in this document, even if considered less likely to have consequences for the scientific benefit-risk evaluation, represent a violation of basic ethical principles in clinical research. Other examples are failure to obtain informed consent or failure to comply with SAE reporting timelines. It is outside the scope of this document to further discuss the nature and extent of such findings that would constitute ethical reasons for rejecting an application.

It should be noted that extensive non-compliance with ethical principles may indicate more widespread problems also affecting aspects of direct relevance to the benefit-risk assessment. But even if not, they should have consequences for the final conclusion on approvability of the application.

2. General note on findings

The GCP inspection findings in the table below are common findings or findings that are considered illustrative, but it should not be seen as an exhaustive list.

Many of the findings (especially in the intermediate “may influence” category) result in increased variability/reduced measurement precision in the assessment of efficacy endpoints. They tend to blur real differences between treatment groups in randomized clinical trials. Sponsors aim to reduce variability by different measures in order to increase the likelihood of separation, for example between study drug and placebo. However in superiority studies, once the study has been completed, and superiority has been established, inspection findings merely indicating increased variability and not introducing bias favouring one treatment over the other are relatively unproblematic in the interpretation of the study results. For non-inferiority studies, the impact of these inspection findings is not as straightforward. Increased variability may disguise a real difference between a superior active comparator and an inferior investigational drug. On the other hand, increased variability tends to widen the confidence interval for the mean difference/ratio between the investigational drug and the active comparator making the non-inferiority claim more difficult to obtain.

Further, it is important to assess whether the findings affect the interpretation of the primary efficacy endpoint or important safety endpoints. Needless to state, the findings have less significance for the

benefit-risk assessment if they only have consequences for secondary or exploratory endpoints. However, it should be born in mind that secondary endpoints may serve important purposes of ensuring internal consistency, in particular in applications with a single pivotal trial.

Finally, even if individual findings in the intermediate and low-impact category may not affect the benefit-risk assessment looked upon in isolation, the combination of several of these findings is an indicator of overall poor data quality and therefore likely to become significant.

3. Categorisation of findings

Inspection findings which are likely to influence the benefit-risk evaluation

1	**Deficiencies in blinding of study medication**	Problems associated with the intentional or accidental loss of blinding of study medication can potentially lead to bias both in terms of interpretation of efficacy and safety/tolerability, for example bias favouring the investigational drug at the cost of placebo or reference medication. The problems are less worrisome if the unblinding was confined to few patients/sites as compared to more systemic faults potentially affecting the entire study. Depending on the circumstances, inappropriate unblinding may be considered less problematic if the study staff unblinded to treatment was not responsible for the evaluation of the primary efficacy endpoint and other important endpoints. Furthermore, it should also be considered to what degree the evaluation of the important endpoints were prone to bias due to knowledge about treatment allocation: Objective endpoints such as survival and biochemical markers as well as endpoints that can be independently adjudicated are likely to be less sensitive to bias than patient reported outcomes, psychometric tests and other endpoints with a substantial subjective component.
2	**Deficiencies in randomization**	All findings suggesting that the treatment allocation in a clinical study intended to be a randomized study was either not truly random or breached after direct or indirect unblinding (see above) will raise questions about the comparability of the treatment groups and the causal relationship between treatment and effect and will consequently have implications for the benefit-risk evaluation.
3	**Violation of diagnostic inclusion- and exclusion criteria**	Deviations from eligibility criteria related to the proper diagnosis of patients raise serious questions whether the patients suffer from the targeted disease. Therefore, deviations will have an impact on the indication that the trial could support and consequently on the benefit-risk assessment. Also, deviations regarding sub-staging and assessment of severity should be considered significant.
4	**Violation of procedures related to the assessment of the primary efficacy endpoint**	Non-compliance with regard to evaluation of important efficacy endpoints will in many cases affect the benefit-risk assessment, especially if a potential bias favouring the test product is introduced or if the clinical entity being measured is significantly changed.

5	**Inadequate reporting of adverse events and other safety endpoints**	Inspection findings indicating systematic underreporting of adverse events at investigational sites or as a more general phenomenon in certain countries/regions participating in a study can seriously jeopardize the evaluation of safety and tolerability and as such influence the overall benefit-risk evaluation. Also inadequate reporting of other safety endpoints will in many cases make the benefit-risk assessment more difficult. The findings may be considered less significant in situations where the safety profile of the drug is well-established, for example in extension applications where the safety and tolerability profile in the new indication is not expected to be significantly different from that of the original indication and patient population.
6	**Missing source documentation**	The ability to verify clinical trial data against source data is considered a key element in GCP. Missing source data which are extensive or which concern diagnosis, primary efficacy assessments and important safety information will have consequences for the assessment of benefit-risk. Shortcomings regarding for example demographic data, other baseline characteristics not related to diagnosis, or secondary/exploratory efficacy endpoints may not be as grave and should be considered individually.
7	**Faults in data management, statistical programming and analyses**	Systemic deficiencies at the level of sponsor/contract research organizations in the set-up of data management, data handling and in the statistical programming of the data analyses (e.g. SAS programming errors) are often difficult to detect at inspections, but have the potential to lead to completely false study conclusions. In other cases, dependent on their nature, they may be less important.
8	**Fraud and other scientific misconduct**	In general, clinical data generated by investigators exhibiting fraudulent behaviour or other scientific misconduct are not reliable and should be excluded from the primary analyses of the study. The impact on the benefit-risk evaluation depends on the relative contribution of the concerned investigators and on the extent to which the failure by the sponsor to detect the fraudulent behaviour can be attributed to deficiencies in the sponsor's quality system. Scientific misconduct at a systemic level (such as data management, statistical analyses and reporting) will certainly have a significant impact on the benefit-risk assessment.
Inspection findings which may influence the benefit-risk evaluation		
9	**Violation of inclusion and exclusion criteria (other than diagnostic criteria)**	Inclusion and in particular exclusion criteria often serve as safety precautions. Failure to meet these criteria may not have consequences for the benefit-risk assessment. Other criteria are introduced to study protocols by sponsors to reduce variability in the efficacy measurement, such as the exclusion of patients with concomitant diseases or taking certain other medications to minimize "noise" potentially affecting the efficacy endpoints. Furthermore, some criteria may be included in the study protocol for no obvious scientific/methodological reason, despite the requirement for clinical studies to be scientifically sound, because it was generic requirement in the company SOP. In the framework of benefit-risk evaluation when assessing a positive clinical study, non-fulfilment of these criteria may not be of major

		importance, at least in superiority trials.
10	**Violation of study procedures regarding rescue medication**	Failure to comply with restrictions on rescue medication can potentially favour the less effective treatment in a randomized clinical trial. In a trial where the test product has been shown to be superior to placebo, this is often not a major concern when interpreting the results. An exception would be superiority trials where reduced use of another medicinal product is an important efficacy endpoint (e.g. reduction of opioid rescue medication in trials of analgesics). In non-inferiority trials, inappropriate use of rescue medication could result in a bias favouring a less effective test product over a more effective active comparator.
11	**Deviations from protocol-specified visit windows**	If these deviations are minor, the findings often reflect too narrow visit windows specified in the protocol. However, the consequences for the understanding of the study results and the benefit-risk assessment can be minor.
12	**Inadequate calibration of instruments, measurement equipment etc. related to the assessment of efficacy**	These findings represent a heterogeneous group. The findings rarely lead to bias favouring the test product. In superiority trials, where superiority of the test product has been proven, the interpretation of the study often remains unaffected.
13	**Rounding issues**	“Rounding issues” refer to non-compliant rounding of numbers (for example 7.31 to 7.4 instead of 7.3). These findings do not often lead to bias favouring the test product. In superiority trials, where superiority of the test product has been proven, the interpretation of the study often remains unaffected.
14	**Failure to document pre-specification of analyses prior to breaking study blind**	It is often stressed that statistical analyses of a clinical trial should be pre-specified as much as possible, preferably before the initiation of the study, but at least before breaking the study blind following completion of double-blind treatment. The reason for this requirement is to avoid analyses which are “tailored” to the outcome of the clinical data, also known as “fishing expeditions”, for example to obtain results favouring the test drug. Furthermore, the unadjusted selection of one analysis method among several ones on the basis of the study results leads to an inflation of type-1 error of the corresponding statistical test and invalidate the treatment comparison irrespective of the validity of each single method itself. For this reason, failure to document pre-specification of analyses will in many cases affect the benefit-risk assessment, in particular when it concerns the primary or other important endpoints. However, there may also be cases where the impact is less. This could be cases where the lack of pre-specification only concerns secondary endpoints, or when guidelines have changed policy regarding efficacy endpoints, or cases where subsequent sensitivity analyses or other analyses defined independent of the unblinded dataset have supported the robustness of the primary analyses first provided by the company.

15	**Discrepancies between the clinical study report and the actual conduct of the study**	This very broad category encompasses findings where the submitted dossier (clinical summary documents and clinical study reports) may well be reflecting what was planned per protocol, but false, misleading or in other ways misrepresenting in terms of what really happened in the study. This would typically be seen as inadequate listing of protocol deviations. The impact on the benefit-risk assessment will depend on the nature of the discrepancies and the extent to which these discrepancies potentially affected critical aspects of the study: Proper diagnosis of patients, primary efficacy endpoints, adverse event recording etc.
16	**Deficiencies in drug accountability**	Records of drug accountability are considered a basic requirement in GCP. They are also often used to assess patient compliance in clinical studies, either as the only assessment or together with other instruments. If other instruments are used (e.g. monitoring of drug levels, patient diaries), the influence on the assessment of compliance may be less. Drug accountability deficiencies can be important for understanding a negative/failed trial, but in many situations they are unlikely to introduce bias favouring the test product. However, they may be important if it is suspected that compliance is significantly poorer in one particular treatment group.
17	**Deficiencies in storage of study medication**	Inappropriate storage of study medication can lead to loss of pharmacological activity of active medicines and may in certain cases also represent a safety hazard to patients (emerging from toxic degradation products), depending on its sensitivity to changes in temperature, light and humidity. In such cases, the finding will have consequences for the evaluation of the sponsor's ability to protect the safety and well-being of trial subjects. But it is not given that the findings will introduce a bias in favour of the tested product and thus have consequences for the benefit-risk evaluation. However, there could be situations where this could be the case, for example non-inferiority studies and studies where comparator medication is affected.
18	**Deficiencies in preparation and administration of study medication at investigational sites**	These findings include problems related to the proper reconstitution of medicines for injection/infusion, use of appropriate containers, equipment and devices for appropriate dosing of study subjects etc. It is not given that such deficiencies will introduce a bias favouring the test product in a superiority trial, but the impact on benefit-risk will have to be assessed in each individual case.
Inspection findings which are less likely to influence the benefit-risk evaluation		
The findings in this category are a heterogeneous group of non-compliance issues related to the conduct of the study that have no or minor direct influence on the benefit-risk assessment. Some of the findings will be trivial in any respect, but others could be critical in the overall evaluation of the clinical development programme and as such be important for the approvability of the application.		

4. Other considerations

How to extrapolate GCP inspection findings

When inspectors identify non-compliance at a particular investigational site with potential impact on the benefit-risk balance, the question arises: Was this finding confined to just this one investigator, or could it be an indication of a more systemic fault in the trial? There is no easy answer to this question, and each case will have to be looked at individually. In general, if the same finding is identified at two different investigational sites (preferably in different countries/regions), it greatly increases the probability that the finding reflects a general problem with the study. As GCP inspections are normally confined to 2-3 investigational sites, which is normally a very limited number compared to the number of sites included in the study, other means may be needed to demonstrate that non-inspected sites are GCP compliant or to quantify the impact if a substantial number of non-inspected sites were to have the same problems as the inspected site(s).

Sensitivity analyses excluding just one or two inspected sites are often not meaningful since these do not address the uncertainty regarding the non-inspected sites. If it is suspected that the GCP inspection finding is likely to be a more systemic problem, but still limited to for example a certain country or region, a specific assessment tool or instrument not used at all sites, or patients experiencing a specific event, sensitivity analyses excluding the affected patients may be helpful in the benefit-risk evaluation.

Sensitivity analyses aimed at investigating the maximum impact of a finding observed at one site on the study results assuming that the finding would have observed at all sites if inspected may also be useful.

If the problem causing an inspection finding leaves an identifiable “footprint” in the clinical database, targeted analyses of the database may be another way to shed more light on the issue of how to generalise the inspection findings. For example, investigators systematically applying identical scores of a psychometric test to patients throughout a study (despite such a scoring pattern being very unlikely in the course of the disease) could be identified by carefully designed searches in the clinical database.

Findings related to inspections of the sponsor/CRO

GCP findings may not pertain specifically to the investigational sites, but to the sponsor, CRO or other organisations or institutions to which responsibilities have been delegated by the sponsor (e.g. the central laboratory). These findings are almost by definition always systemic, and some examples are given above: Deficiencies in blinding of study medication; deficiencies in randomization; faults in data management, statistical programming and analyses; failure to document pre-specification of analyses prior to breaking study blind; discrepancies between the clinical study report and the actual conduct of the study. Other findings may point to extensive deficiencies in the overall monitoring and quality system of the sponsor, CRO or affiliated organisations or institutions. These findings question the overall validity of the clinical data and thus likely to have consequences for the benefit-risk evaluation, but each case will have to be evaluated individually. For example, critical deficiencies observed in the quality system of a central laboratory responsible for blood tests of importance for the appropriate diagnosis of patients or the evaluation of primary efficacy endpoints or important safety endpoints should have impact on the benefit-risk assessment.