EUROPEAN MEDICINES AGENCY
SCIENCE MEDICINES HEALTH

05 January 2019
EMA/CHMP/205/95 Rev.6
Committee for Medicinal Products for Human Use (CHMP)

Guideline on the clinical evaluation of anticancer medicinal products

Draft agreed by Oncology Working Party	24 March 2020
Adopted by CHMP for release for consultation	5 October 2020
Start of public consultation	13 November 2020
End of consultation (deadline for comments)	15 February 2021

This guideline replaces guideline on the evaluation of anticancer medicinal products in man' EMA/CHMP/205/95 Rev 5

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

Keywords	***Cancer, malignancy, biomarker, targeted drugs, pharmacogenomics***

Official address Domenico Scarlattilaan 6 ● 1083 HS Amsterdam ● The Netherlands
Address for visits and deliveries Refer to www.ema.europa.eu/how-to-find-us
Send us a question Go to www.ema.europa.eu/contact **Telephone** +31 (0)88 781 6000
An agency of the European Union

Table of contents

Executive summary

The purpose of this guideline is to provide guidance on all stages of clinical drug development for the treatment of malignancies, including drug resistance modifiers or normal tissue protective compounds. Supportive measures such as anti-emetics and haematopoietic growth factors, however, are covered by separate guidelines.

Alongside conventional aims such as defining the proper dose(s) and schedule(s), the importance of identifying a target population with optimised benefit risk is emphasised in Section 6: Exploratory Studies. Guidance is also provided on combination studies. Combinations of drugs with minimal activity as monotherapy, but synergistic effects when combined, as well as combinations of conventional cytotoxics, are also discussed.

Section 7 discusses the design of confirmatory and the choice of endpoints. Convincingly demonstrated favourable effects on overall survival (OS) are from both a clinical and methodological perspective the most persuasive outcome of a clinical trial aiming to demonstrate efficacy. Other possible primary efficacy endpoints include progression-free or disease-free survival (PFS/DFS), and patient-reported outcomes.

An assessment of benefit/risk should encompass all relevant data on efficacy and safety, also taking into account uncertainties as well as external data of relevance to the experimental compound and the disease to be treated.

The requirements of the characterisation of the safety profile have changed with the emergence of molecularly targeted agents (MTAs), immunomodulating drugs and other non-cytotoxic agents. These types of agents may have other types of toxicity and are often dosed differently compared to conventional chemotherapy. The dose-finding process and concepts such as dose limiting toxicity (DLT) may therefore need to be addressed differently than for standard cytotoxic agents. This is discussed in Section 6.2.1. Furthermore, cumulative incidences of adverse events by toxicity grade alone are not sufficient to characterise the toxicity profile. The impact of an adverse drug reaction (ADR) on the benefit-risk balance may for example differ importantly depending on how the incidence, prevalence and severity change with time on treatment, and on the possibility to alleviate the ADR by dose reduction or interruption. This is addressed in Section 9.

Definitions and abbreviations used in this guideline are summarised at the end of the document. Appendix 1 provides methodological guidance on the use of progression-free survival (PFS) as endpoint in confirmatory studies. Appendix 2 focuses on the use of patient reported outcome (PRO) measures

and health related quality of life (HRQoL) from a regulatory perspective. A revised paediatric guideline is also foreseen as Appendix 3, and Appendix 4 is dedicated to condition-specific guidance.

1. Background

The guideline on anticancer medicinal products adopted in 1996, and revised in 2001 and 2003, focused on conventional cytotoxic compounds. In 2005, a major revision was undertaken, aiming at covering non-cytotoxic compounds, to expand on the sections on exploratory trials and to provide more guidance with respect to methodological issues. Later, an appendix on methodological issues related to the use of PFS was added and in early 2010 an appendix on haematological malignancies followed. The main guideline was subsequently updated in line with these appendices, e.g. with regard to confirmatory studies based on aims of therapy and relative toxicity, while the section on condition specific guidance was expanded and placed in a separate Appendix 4.

Since then a new Appendix 2 has been adopted, concerned with patient reported outcomes and health-related quality of life.

The purpose of the 5th revision of the main guideline is to address current changes in the therapeutic landscape that affect the requirements with regard to collection and reporting of safety data in order to inform the benefit-risk evaluation, including a need for more differentiated and detailed safety data presentation.

This 6th revision addresses the most recent designs in oncology (such as umbrella and basket trials, so-called master protocols) and the emergence of indications defined in the first place by a biomarker selective for a disease sensitive to the treatment.

2. Scope

Whilst the thrust of a regulatory guideline should be on confirmatory studies, the aim of this guideline is also to underline the use of exploratory studies in order to identify the most appropriate target population in addition to the usual aims: to define dose, schedule, tumour type and line of therapy. The role of biomarkers to achieve these objectives is also further emphasised in this revised guideline.

There are numerous possible ways to classify anti-cancer drugs such as direct anti-tumoural vs. indirect anti-tumoural, or based on pharmacology or molecular target (e.g. hormones, immune modulators, nuclear-targeting, signal-transduction targeting, etc.). As this document is meant to provide guidance on clinical drug development, the aim has been to classify compounds according to reasonable designs of exploratory studies, i.e. cytotoxic compounds where toxicity and objective response rate (ORR) are considered suitable markers of activity in dose finding studies vs. non-cytotoxic compounds where ORR and/or toxicity may not serve this purpose.

A very large number of anti-cancer compounds have been and currently are under development. Only a minority, however, have completed the clinical development and obtained a marketing authorisation, due to poor activity or evidence of a detrimental safety profile. Until non-clinical models with good predictive properties have been defined, this situation is likely to remain essentially unchanged and the absence of such models is considered to constitute the greatest hurdle for efficient drug development within the foreseeable future.

Since chemoprotective agents and drug resistance modifiers are used as part of anticancer regimens, some guidance on these agents will also be provided in appropriate sections of this guideline. Anti-emetics and haematopoietic growth factors, however, are covered in separate documents.

3. Legal basis

This document should be read in conjunction with Directive 2001/83/EC, as amended. Applicants should also refer to other relevant European and ICH guidelines on the conduct of clinical trials, including those on:

- Nonclinical evaluation for anticancer pharmaceuticals EMEA/CHMP/ICH/646107/2008 (ICH S9)
- Clinical Investigation of the Pharmacokinetics of Therapeutic Proteins CHMP/EWP/89249/2004
- Evaluation of the Pharmacokinetics of Medicinal Products in Patients with Impaired Hepatic Function - CPMP/EWP/2339/02
- Guideline on the investigation of drug interactions, CPMP/EWP/560/95/Rev. 1
- Points to Consider on Adjustment for Baseline Covariates - CPMP/EWP/2863/99
- Points to Consider on Multiplicity Issues in Clinical Trials - CPMP/EWP/908/99
- Guideline on the choice of non-inferiority margin - CPMP/EWP/2158/99
- Qualification of novel methodologies for drug development: guidance to applicants EMA/CHMP/SAWP/72894/2008 Rev.1
- Guideline on clinical trials in small populations-CPMP/EWP/83561/2005
- Choice of Control Group in Clinical Trials CHMP/ICH/364/96 (ICH E10)
- Guideline on clinical evaluation of diagnostic agents - CPMP/EWP/1119/98
- Note for guidance on clinical safety data management: data elements for transmission of individual case safety reports - CPMP/ICH/287/95 (ICH E2B)
- Points to consider on application with 1. Meta-analyses 2. One pivotal study - CPMP/EWP/2330/99
- Reflection paper on methodological issues in confirmatory trials planned with an adaptive design – CHMP/EWP/2459/02
- Guideline on the investigation of subgroups in confirmatory clinical trials - EMA/CHMP/539146/2013 adopted 31.01.2019
- Guideline on strategies to identify and mitigate risks for first-in-human and early clinical trials EMEA/CHMP/SWP/28367/07 Rev. 1
- Addendum on terms and concepts of pharmacogenomic features related to metabolism to the Guideline on the use of pharmacogenetic methodologies in the pharmacokinetic evaluation of medicinal products EMA/CHMP/37646/2009
- Reflection paper on Pharmacogenomics in oncology EMEA/CHMP/PGxWP/128435/2006
- Reflection paper on Methodological issues with pharmacogenomic biomarkers in relation to clinical development and patient selection EMA/446337/2011,
- Guideline on Good Pharmacogenomic practice - EMA/CHMP/718998/2016

- Guideline on key aspects for the use of pharmacogenomic methodologies in the pharmacovigilance evaluation of medicinal products EMA/CHMP/281371/2013).
- Reflection paper on methodological issues in confirmatory clinical trials planned with an adaptive design CHMP/EWP/2459/02)

Guidance on specific aspects of paediatric medicinal product development is available in:

- -Guideline on pharmaceutical development of medicines for paediatric use (EMA/CHMP/QWP/805880/2012 Rev. 2)
- -Guideline on the role of pharmacokinetics in the development of medicinal products in the paediatric population (EMEA/CHMP/EWP/147013/2004)
- -Reflection paper on the use of extrapolation in the development of medicines for paediatrics (EMA/199678/2016)

- -Draft guideline on good pharmacovigilance practices (GVP) - Product- or population-specific considerations IV: paediatric population (EMA/572054/2016).
- Lastly, some additional considerations that apply exclusively to the paediatric population are discussed in the Note for guidance on evaluation of anticancer medicinal products on man: Addendum on paediatric oncology (EMEA/CPMP/EWP/569/02, under revision).

4. Pharmacokinetics

In general, the same recommendations are valid for anticancer products as for other medicinal products and reference is made to the clinical pharmacology guidelines available. For therapeutic proteins, reference is made to CHMP/EWP/89249/2004. This section is thus mainly meant to highlight some areas where missing information frequently has been encountered in submissions for marketing authorisation and to underline some areas considered to be of special interest.

In the past, human mass-balance studies (in vivo studies investigating the fate of a radiolabelled dose in plasma and excreta) have not been performed to the same extent for anticancer drugs as for other medicinal products. Due to the importance of the information gained in these studies for the understanding of the clinical pharmacology of the investigational drug, including the drug-drug interactions assessment, mass-balance studies are strongly recommended (CPMP/EWP/560/95/Rev. 1).

Food interaction studies should be performed prior to phase III and administration in fed or fasted state should be investigated and a rationale for administration in fed and/or fasted state should be provided.

The potential for drug-drug interactions should be assessed. If in vitro data indicate that the anticancer product will give rise to, or be a victim of, important drug-interactions, this should as far as possible be investigated in vivo.

Studies to be undertaken in patients with impaired organ function should mainly be selected based on prior information on the mode of elimination of the drug and formation/elimination of potential pharmacologically active metabolites. If a study in hepatic impairment is needed and liver metastases are common in the target patient population, as a first step a study in patients with liver metastases is warranted. Whether studies in more advanced liver disease are needed should be decided on a case by case basis (CPMP/EWP/2339/02). Lack of data is reflected in the product information, i.e. Summary of product characteristic (SmPC). Exploratory studies, including PK, in patients with malignant ascites or other third space conditions such as massive pleura fluid are encouraged if seen in the condition being treated.

It is recommended to also evaluate the influence of intrinsic factors through population PK analyses. The plasma concentration data should optimally come from as many as possible of the clinical studies. Both sparse (few samples per patient) and rich data (full plasma concentration-time profiles) can be used. Factors to investigate as covariates could include age, weight, gender, renal function, S-bilirubin, liver enzymes, genotype, soluble receptors/ligands, tumour burden, inflammatory markers etc.

The use of PK and PD (biomarkers and clinical markers) sampling for PK/PD analysis related to efficacy and safety is encouraged. This information aids in understanding the exposure-response relationships for the drug, and may allow for a rational selection of treatment strategies in patients who are at risk for excessive toxicity or ineffective therapy. Exposure-efficacy and exposure-safety analysis/modelling is encouraged in the Phase II randomized trials (sections 6.2 and 6.3) to provide PK/PD information and to support Phase III dose selection. Ultimately, a pooled analysis of PK and PD data obtained in all phases of development is encouraged in order to fully characterize and summarize the PK/PD of the drug. In order to utilize all collected data efficiently, longitudinal PK/PD analysis of PD data e.g. tumour shrinkage as a continuous variable is recommended. Simulation based evaluations of the study design with respect to power of identifying PK/PD relationships and covariate effects are recommended. Due

to high withdrawal rates leading to informative censoring, handling of missing data is of crucial importance in longitudinal analyses and sensitivity analyses, e.g. using early time points for tumour shrinkage should be considered.

5. Biomarkers

Biomarker investigations in the context of regulatory submissions should be accompanied by a full description of: the nature and functional role of the biomarker, the hypothesis regarding the relationship between the biomarker and the drug's effects, the purpose and intended context of use, the analytic method by which and the source/matrix of tissue/biomaterial in which the biomarker is measured, and the analytical and clinical performance characteristics. The use of available scientific guidelines on reporting results of biomarker analyses is also encouraged in order to facilitate uniform reporting and assessment of results.

Biomarker investigations, either for exploratory or confirmatory purposes, are a crucial element of anticancer drug development. Biomarkers can serve a wide spectrum of purposes, including establishing early proof of concept, determining the optimal biological dose (section 6.2.1), identifying response/resistance mechanisms, prospectively selecting patients for treatment, assessing/monitoring efficacy and safety, and guiding posology. Apart from using some biomarkers as a surrogate endpoint to clinical outcome (see below), biomarkers are primarily used to either characterize patients with respect to a specific disease *prognosis* or to identify patients that are expected to benefit from a given treatment more than others. Whereas the first is referred to as *prognostic*, the latter one is designated as *predictive*. In drug development, the main focus lies on *predictive* biomarkers intended to determine the best treatment option for a specific patient.

Sample collection

The clinical studies performed in the context of obtaining marketing authorisation are the key opportunity to gather tumour tissue and other biomaterial for biomarker analyses. While collection of tissue should always be considered in light of associated patient burden, it is generally considered reasonable to expect that tumour tissue for biomarker analyses is collected at all stages of the development trajectory. It is recommended to collect and store tumour tissue samples suitable for the different types of analyses that can be anticipated (e.g. both fresh-frozen tumour tissue and formalin-fixed tumour tissue). The general principles of collection, processing, transport, storage, and disposition of samples should be adhered to in order to assure sample quality. The general principles outlined in ICH E18 of collection, processing, transport, storage, and disposition of samples should be adhered to in order to assure sample quality.

The source and quality of the tissue samples should be appropriately justified in relation to the type and purpose of the biomarker analysis. Archival tumour tissue samples may not always be suitable for biomarker analyses performed in confirmatory studies, because they are usually obtained under variable conditions (leading to uncertain sample quality) and the time between collection of tissue and the moment at which the patient starts study treatment can vary widely. Freshly obtained tissue collected using standardised procedures for collection and sample processing will generally be preferred.

Baseline tumour tissue collection and analysis is crucial for the investigation of the impact and value of biomarkers. The timing of baseline biopsies should generally be close to the start of study treatment (i.e. during the screening phase), taking into account a wash-out period after prior treatment if appropriate. Additional aspects related to the source of tissue and timing of sampling that should be considered are variability in expression of the biomarker within tumour lesions and/or between tumour lesions in the same patient, which may have impact on the ultimate performance of the biomarker, and temporal variation in biomarker expression, e.g. with tumour progression or in relation to biological cyclic activities.

The collection of on-treatment biopsies should be considered, in particular in early proof of concept studies, e.g. to determine whether the drug modulates its target in tumour tissue. When it is of value to characterise secondary resistance mechanisms, the collection of tumour tissue at the time of progressive disease should be considered.

The collection of circulating tumour DNA (ctDNA; also referred to as free tumour DNA, ftDNA) or circulating tumour cells (CTCs) from blood samples, often referred to as liquid biopsies, are alternative/complementary technologies that allow easy and repeated sampling, e.g. for patient selection, monitoring of drug response or monitoring of development of resistant clones. When such technologies are used, they should be justified based on correlational analyses between tumour DNA and ctDNA or CTCs, in particular in case ctDNA or CTC is intended to be used as a surrogate for mutations present in tumour lesions.

Samples for pharmacogenomic evaluation in relation to pharmacokinetics, safety issues, etc., should be collected and analysed as appropriate.

Biomarker investigations in confirmatory studies

The methodological considerations in relation to biomarker investigations in confirmatory studies are extensively addressed elsewhere (EMA/446337/2011), but several key aspects relevant to anticancer drug development are highlighted here.

Upfront planning of biomarker investigations

The role of the biomarker and biomarker-related hypotheses should be defined upfront as much as possible. For any biomarker, biological plausibility should be discussed, but some clinical data are generally needed to substantiate clinical relevance.

If the biomarker is already well developed, cut-off points have been defined as appropriate, and the predictive ability of the biomarker can be estimated, it will be possible to stratify patients in phase III trials (and in some cases phase II trials) according to biomarker status or different cut-offs and confirm and validate the predictive ability of the biomarker.

In many other cases where a rationale for the biomarker is present but the current knowledge is insufficient to aim for confirmation, strengthening the design and analysis of the biomarker investigations should be considered at the study design stage. This could include sample size/power calculations to ensure that sufficient information across the range of biomarker values/cut-offs is available, and preplanning of external or internal validation of the subgroup results, e.g. by using cross-validation approaches. Ideally, replication of findings in a set of two pivotal clinical trials should be planned. Scientific advice regarding the planned biomarker development strategy is strongly recommended.

Ensuring a representative biomarker-evaluable population

It is generally preferred that sampling of biomarkers is planned to be complete (e.g. by requiring biopsy as an inclusion criterion in the confirmatory study), if appropriate. Pre-planned sampling of a representative subgroup may be possible under certain conditions (e.g. based on a random selection mechanism), but this will reduce statistical power and lead to less precise estimates of treatment effects. Reasons for lack of availability of samples should be recorded. Potential selection bias and non-random missing data should be investigated, as appropriate.

Subgroup investigations for predictive biomarkers in confirmatory studies

In many cases, candidate predictive biomarkers have been identified prior to initiation of the confirmatory phase III clinical studies. Examples are markers that may affect efficacy (e.g. presence/absence of different driver mutations or primary resistance mechanisms) and/or safety (e.g. genetic polymorphism in drug-metabolising enzymes). Subgroup analyses should in this case be pre-planned, adjusting for multiplicity, as appropriate, to assess the clinical consequences of these factors. When a drug is being developed in a disease setting where other available/approved treatments are administered based on biomarkers (e.g. driver mutations), it is recommended to determine these

biomarkers also in the context of the clinical studies of the drug under investigation (for example, for a new drug developed in melanoma, determination of *BRAF* mutation status should be considered).

Biomarkers for upfront patient selection in confirmatory studies (enrichment)

When a biomarker is used to select patients for treatment – i.e. the biomarker is used to enrich the study population and to define the target population accordingly – the predictive value of the biomarker should be established. This will normally require at least a limited amount of clinical data in the biomarker negative population.

If the biomarker used for patient selection is essentially a continuous marker (e.g. different degrees of expression or mutation counts) and a cut-off is used to classify patients as biomarker-positive or negative, thorough justification of the adequacy of the cut-off value is required. Furthermore, cut-off values should be defined *a priori* (e.g. based on prior data) and validated in the confirmatory clinical studies. When patients are selected upfront based on a continuous marker, it is also important to perform pre-planned subgroup analyses assessing the association between degree of marker expression and outcome within the population enrolled.
In case evolving knowledge from the phase III clinical trials suggests that the cut-point may need to be refined, availability of independent data from a second clinical trial to validate the usefulness of the change in definition is crucial.

While (enrichment) biomarkers used to select patients for treatment can be purely prognostic (providing information about the patient's overall disease outcome) or predictive (providing information about the effect of a therapeutic intervention), many of the biomarkers that are considered predictive are also prognostic (e.g. HER2 expression in breast cancer). In some cases, the prognostic association will be relatively well characterised, but for many novel markers this is often not the case (consider e.g. PD-L1 expression). The unknown prognostic effect particularly underscores the need for controlled data with an adequate comparator when the confirmatory study is performed in a biomarker-enriched population, in order to be able to adequately determine the drug's treatment effect and distinguish it from any prognostic effects. If the marker is prognostic and/or predictive, a stratified analysis for the degree of marker positivity should be foreseen.

For targeted agents that are used in enriched patient populations where the biomarker used to select patients comprises more than one entity (e.g. different mutations in the same gene carrying potentially different predictive information, such as KIT and/or PDGFRα mutations in GIST), the patient selection strategy in the confirmatory studies should be adequately justified, based on available non-clinical/translational data, biological rationale, and supported by available clinical data – aiming at maximising exposure to/treatment of those subsets expected to benefit. In situations where specific (low-prevalence) variants of the marker are present, enrolment of a minimum number of patients carrying the uncommon variants of the marker may be needed for pre-planned subgroup analyses.
When an anticancer medicinal product is developed for use in special populations (e.g., paediatric patients), the clinical validity of the biomarker may need to be established specifically for that population (e.g., validation studies, extrapolation).

Biomarkers as clinical trial endpoints

As a distinct context of use, some biomarkers are used as clinical trial endpoints. However, for acceptance of these biomarkers as a surrogate endpoint used to support benefit/risk assessment in a regulatory submission, it is crucial that clinical validity is comprehensively established regarding the relationship with a treatment effect in the clinical endpoint, in addition to analytical validity (see below) *prior* to use in confirmatory studies. For new, non-established endpoints, requesting scientific advice regarding their use or qualification is always recommended.

Biomarker assays

Any biomarker assay used in the context of anticancer drug development should be substantiated by data supporting its analytical validity, which needs to be adequate considering the context of use of the biomarker/assay. It is acknowledged that biomarkers measured in early clinical trials are often

more exploratory in nature than those used in later stages, but it is essential that also for these biomarker assays analytical validity is sufficiently assured (EMA/CHMP/SAWP/72894/2008 Rev.1, EMEA/CHMP/PGxWP/128435/2006, EMA/CHMP/641298/2008). This is, for example, particularly relevant for biomarker assays used in early clinical studies to select patients/determine eligibility for study treatment. Changes in clinical trial assays between different clinical phases of the drug development programme should be minimised as much as possible. In cases where there were changes to clinical trial assays were performed or where more than one assay was used during the development, evidence of concordance should be provided.

Centralised testing to determine biomarker status is recommended for confirmatory/pivotal studies, while local testing could be considered as a secondary analysis. For simple assays, local testing alone may be sufficient if assay standardisation can be assured. Analysis of concordance between central and local assessment of biomarker status may be useful to gain insight into performance of the assay in the setting of routine clinical practice.

In cases where the identification of the biomarker is essential for the safe and effective use of a medicinal product, co-development (or close-knit development) of the diagnostic assay and the medicinal product is encouraged.

6. Exploratory studies

Exploratory studies are essential in rational drug development. The distinction between Phase I/II exploratory and Phase III confirmatory trials has been adhered to in this Guideline. However, this does not mean that exploratory aims should not form an important part of Phase III trials. Similarly, hypothesis generation, testing and confirmation may form parts of Phase II trials.

So called phase 0 trials, i.e. trials exploring micro dosages may be informative in certain circumstances as regards tissue distribution and receptor binding, e.g. when it is considered important to early identify whether a compound is likely to penetrate different anatomical or physiological compartments, such as the central nervous system (CNS), or, when feasible, to obtain early data on pharmacological activity at low drug concentrations.

6.1. Cytotoxic compounds

This section refers to conventional cytotoxic agents, i.e. compounds inducing irreversible lethal cellular damage following short-term exposure through interference with DNA replication, mitosis, etc. For these compounds, toxicity and tumour response are considered suitable indicators of activity.

Conceptually this section is also relevant to more targeted cytotoxic compounds such as monoclonal antibody coupled toxin products. In these circumstances however, tumour antigen expression and prodrug activating pathways should also be taken into consideration.

As for non-cytotoxic compounds, non-clinical and clinical studies encompassing aims to characterise prerequisites for activity/resistance and to identify markers of resistance are encouraged.

6.1.1. Phase I, single agent dose and schedule finding trials

The basic assumption governing the design of these trials is that, for dose finding purposes, toxicity is an acceptable endpoint. The main objective is thus to define dose-limiting toxicities and the dose to bring forward into further trials. While meeting this objective is generally straightforward, in spite of the fact that the inter-patient variability in PK might be large, it is often more complex to define reasonable dose schedules to study further.

Initial dosing may use flat doses or body surface area (BSA) scaled doses. The scientific support for the notion that BSA scaled dosing generally reduces inter-patient variability in exposure is weak and may lead to over and under-exposure in patients with a high and low BSA, respectively. It is expected that the importance of BSA or weight for variability in exposure is explored through modelling & simulation using actual pharmacokinetic data.

The use of pharmacodynamic endpoints, where available, may also assist in dose selection.

Main objectives

- Maximum Tolerated Dose (MTD), Dose Limiting Toxicity (DLT) and recommended Phase II dose (RP2D) should be identified for defined schedules and modes of administration.
- Frequent side effects and target organs for toxicity should be characterised as regards relationship to dose and schedule. Severity, duration and reversibility should be determined.
- Initial characterisation of pharmacokinetics including dose and time-dependencies. As appropriate, PK/PD related to target effects and adverse effects, exposures obtained with different routes of administration.

Eligibility of patients

These trials should normally be undertaken in cancer patients without established therapeutic alternatives.

Routes of administration and schedules

The choice of route and rate of administration of the first dose in man should be justified based on the non-clinical data. In most cases, intravenous administration, when feasible, is advisable for first use in man studies since it eliminates variability related to bioavailability.

For schedule finding, experience related to class of compounds is helpful. Non-clinical data with respect to cycle dependency and the ratio tumour / normal tissue cytotoxicity *ex vivo* may be of some interest.

Dose escalation

In case of minimal toxicity, or occasionally in case of non-significant toxicity, within-patient dose escalation may be appropriate in order to reduce the number of patients exposed to non-active doses. This may be acceptable after the end of the period of DLT assessment, if non-clinical data provide evidence of no cumulative toxicity.

If toxicity is acceptable, the patient may be re-exposed upon resolution of toxicity and preferably should receive at least 2 cycles at the same dose level.

Evaluation of toxicity

The minimal requirements for evaluation of adverse effects include assessment of symptoms, physical examination, ECG, blood and urine laboratory analyses and radiological assessment as appropriate. Preclinical data should be used to guide the need for further examinations. If there are no signals with respect to QTc in preclinical studies or related to class of products, no dedicated QTc studies are expected, but inclusion of ECG as part of routine monitoring is recommended. Local toxicity at the site of administration should be specifically recorded. The toxicity should be graded according to a generally recognised system, e.g. the National Cancer Institute's (NCI) Common Terminology Criteria for Adverse Events (CTCAE).

Factors influencing toxicity (organ dysfunction, concomitant therapy) should be explored as appropriate. These factors should be further elucidated in Phase II/III.

6.1.2. Phase II, single agent therapeutic exploratory studies

Phase II trials may investigate single-agent activity in a variety of tumour types, or in a selected tumour type, or investigate activity and feasibility of combination or multimodality regimens.

This section is focused on trials where the primary objective is to estimate single agent anti-tumour activity in patients with a defined tumour type in order to identify compounds to bring forward to confirmatory trial.

Objectives and design

Phase II trials may use a variety of study designs and early studies should provide initial evidence of treatment activity and tolerability. Inclusion of a randomised control arm is encouraged, particularly if only one confirmatory pivotal trial is foreseen (see Section 7.1.2).

The studies are intended to:

- Assess the probability of response (and other relevant efficacy measures) in the target tumour type and conclude on the need for further studies (investigate earlier stages of the disease, combinations, compare with standard therapy).
- Investigate pharmacogenomics and biomarker characteristics, where appropriate.
- Further characterise dose and schedule dependency, with respect to safety and activity.
- Further characterise the side-effects of the medicinal product.
- Further characterise PK and PK/PD (see section 4).
- When applicable, further characterise the optimum route of administration.

Selection of patients

Exact definition of the target disease, previous therapy (if any) and stage should be given, in line with internationally agreed diagnostic criteria.

Provided safety and activity is reasonably established and there is a scientific rationale, it might be appropriate to conduct studies also in patients for whom alternative therapies are available. This includes the neo-adjuvant setting in treatment naïve patients scheduled for surgery, provided that delay in surgery cannot be unfavourable to the patient. The safety and interests of the patient must always be guaranteed, and a detailed justification should be provided in the study protocol. In these cases, the use of sensitive measures of anti-tumour activity such as functional imaging is expected.

Dose and schedule

The dose and schedule should be clearly defined. Details on the administration of the medicinal product with special precautions (hydration of patients, protection against light and temperature, etc.) should be stated as well as other agents, which are contraindicated during the study period.

- Guidance should be supplied outlining dose reductions related to the severity of the observed toxicity.
- As appropriate, guidance outlining dose escalations in case of low toxicity may be incorporated.
- Consideration should be given to study high-risk patients (e.g. high risk with respect to target organ toxicity or compromised metabolic or excretory mechanisms for the experimental compound) separately.
- Any evidence of cumulative toxicity should be recorded and estimated as a function of total dose. This should be specifically studied according to target organ or function.

Evaluation of activity

ORR should be documented according to international standards (e.g. RECIST, Volumetric RECIST or WHO criteria). Modifications of these criteria may be appropriate in certain situations, but should be justified.

In evaluating ORR, the intention-to-treat (ITT) principle should be adhered to. In single arm studies, ORR in the per-protocol analysis set may be reported as primary outcome measure. External independent review of tumour response is encouraged, according to the objectives of the trial.

Data on duration of response, TTP/PFS, confirmed ORR and available data on OS should normally be reported. The use of tumour biomarkers and other dynamic measures of activity is encouraged.

In haematological malignancies, disease specific response criteria are unavoidable in many cases and full harmonization has not yet been accomplished for some disease entities. Therefore it is of importance to follow the progress made by international working groups on these issues. Especially if less conservative disease specific response criteria are introduced in new clinical guidelines, a justification with focus on aspects of drug development is expected from the sponsor.

In patients with symptomatic disease at baseline, the assessment of symptom control is encouraged if a randomised phase II trial is undertaken.

6.2. Non-cytotoxic compounds

This refers to a very heterogeneous group of compounds ranging from antihormonal agents to antisense compounds, signal transduction, angiogenesis or cell cycle inhibitors, immune modulators, etc. The common element affecting the design of clinical trials is that toxicity may not be an appropriate endpoint in dose and schedule finding trials and ORR may not be an appropriate measure of anti-tumour activity.

In contrast to cytotoxic chemotherapy, these compounds are typically administered continuously, and the toxicity profiles tend to differ so that DLTs may occur first after multiple cycles of therapy. This is of importance for the recommended Phase II dose (RP2D) in cases where tolerability and toxicity guide dose selection and may require alternative strategies with regard to definition of DLT and MTD.

For these reasons, the early stages of clinical drug development are more complex and have to be tailored according to the assumed pharmacology of the individual compound as defined in non-clinical studies. The rather strict delineation between Phase I and II trials, as for conventional cytotoxic compounds, may be less relevant as measures of anti-tumour activity, e.g. based on assessment of biomarkers might be needed early in order to define dose and schedule.

Otherwise, most of the elements discussed in relation to cytotoxic drugs are of relevance also here such as restrictions with respect to patient eligibility, recommendations as regards routes of administration, evaluation of toxicity and anti-tumour activity, etc. These issues will not be further discussed here.

6.2.1. Phase I, single agent dose and schedule finding trials

Non-clinical data and, when available, data from healthy volunteers should be used to design the studies to be conducted in patients, e.g. as regards eligibility criteria and starting dose, as well as in terms of agent-specific toxicities to follow and appropriate safety observation time. In accordance with the guidance for cytotoxic compounds, availability of established therapies should normally be regarded as an exclusion criterion. Refractoriness to conventional cytotoxic compounds, however, may confer resistance also to some clearly non-related compounds. This obviously affects the possibility to define a dose/concentration – effect relationship. All sensible and ethically acceptable measures undertaken to increase the assay sensitivity of these clinical trials, including the conduct of window of opportunity studies (Definitions and Abbreviations) are encouraged. Whenever appropriate, this includes measuring the expression of the assumed target(s) for drug activity.

PD measures may include biochemical measures (receptor binding, enzyme inhibition, downstream events, etc. as defined in non-clinical studies), functional imaging, proteomics, immunological measures (antibody or T-cell response), etc. Population PK/PD studies are encouraged. For compounds shown to be cytostatic in non-clinical models, prolonged exposure may be needed to elicit tumour shrinkage in clinical studies. If in these cases unexpected, early tumour shrinkage is observed this

constitutes a signal indicating that further studies exploring the underlying mechanisms behind early response are warranted. While it is acknowledged that drug development for compounds with a single main target for activity, such as mutated BRAF, is more straight forward, it is still expected that the pharmacological rational behind poly-targeting compounds is reflected in the exploratory studies programme, e.g. in terms of biomarkers selected in order to identify the proper target population for treatment. Non-clinical studies should also explore mechanisms of primary or secondary resistance to drugs. This is particularly important for the development of targeted drugs: an identified factor of sensitivity to the drug, crucial to tumour survival/development will normally explain why the drug is active (can induce e.g., shrinkage, slow progression). The elementary mechanism(s) contributing to tumour development is (are) called driver(s). Clonal selection, development of new resistance mechanisms, emergence of a pre-existing alternative driver insensitive to the drug under development may explain why some tumours escape to the drug activity, despite expression of the marker.

Main objectives

- Tolerability, safety, PK and, if at all possible, PD measures of activity are appropriate objectives.
- As for conventional cytotoxic drugs, the use of tumour markers and sensitive imaging techniques, in combination with conventional methods, are recommended in order to delineate possible anti-tumour activity. It is recommended that technical standardisation of, e.g. functional imaging techniques and biomarker assays is implemented in order to reduce inter-centre variability.

Eligibility of patients and methodological considerations

Based on preclinical tolerability and toxicology findings and the assumed pharmacology of the compound, early trials may sometimes be conducted in healthy volunteers.

Eligibility criteria and the number of patients should be defined according to the objectives of the study, also taking into account variability in PK and PD at doses and schedules selected for further studies.

If not pharmacologically justified, proper analyses of biopsies from accessible tumours (primaries and/or metastatic lesions), are expected to constitute a pivotal role in studies undertaken to identify the proper target population for confirmatory studies. This might be crucial and has to be considered in the recruitment of institutions, investigators and patients.

Dose escalation

Until now available experience indicates that tumour selectivity is not to be expected for most compounds. Although dose-safety relationship cannot always be established, tolerability and toxicity remain important measures in dose and schedule finding studies. However, there are cases where dose escalation to MTD is not adequate in order to define the recommended dose. In these cases, dose escalation can be based on pharmacodynamics and safety data in relevant animal models, and on human PK/PD data from initial and subsequent dose cohorts. Mechanism-based PK/PD modelling may also be useful to guide decision making.

In particular in the case of dose-finding for molecularly targeted agents (MTAs), the dose-finding strategy should not only focus on safety endpoints, but also on determining an optimal biologically active dose (alternatively termed "optimal biological dose" or "optimum biologic dose"). This refers to a dose at which optimal biological response according to a predefined effect marker is achieved (e.g. as determined in tumour tissue response) and giving a higher dose does not further improve outcomes (i.e. a dose somewhere at the beginning of the plateau of the dose–response curve). Examples include escalating doses until a target-mediated biologic pathway is optimally altered or escalating doses until a target becomes saturated with the drug, while minimizing the dose required to achieve this maximum pharmacodynamic effect (thereby aiming to minimise toxicity). Preferably a combination of pharmacokinetic/pharmacodynamic endpoints and clinical response endpoints (e.g. objective tumour

response or progression-free survival), in addition to safety endpoints is used to determine the optimal biologically active dose.

Careful consideration must be given to how the concepts of MTD and DLT are pre-defined, in order to capture relevant toxicities and arrive at a useful RP2D.

Many MTAs and immunomodulating therapies will be given continuously/daily (with or without planned off-treatment periods) and/or for prolonged periods of time. Furthermore, certain types of agent-specific toxicity often present after the first treatment cycle, such as peripheral neuropathy from some inhibitors of the ubiquitin-proteasome pathway. Standard definitions for cytotoxic agents, typically focused on acute toxicities in Cycle 1, may therefore not be applicable. Lower grade toxicity over longer periods of time that affect tolerability and the possibility of maintaining the intended dose intensity may need to be addressed in the DLT and MTD definitions.

It has been observed that in phase I trials of MTAs, more than half of the patients present with their first grade 3-4 toxicity after cycle 1. Broader DLT definitions with longer pre-defined DLT/safety observation periods may therefore be relevant to consider. A distinction between cycle 1 acute toxicity, prolonged toxicity impacting on tolerability and late severe toxicity may be informative. Dose escalation based on first cycle adverse events (AEs) may still be reasonable thereby balancing the need to rapidly achieve active dose intensity and the possible need for later dose reductions. AEs should therefore always be reported by treatment cycle and the RP2D should be based on an integrated assessment of likely adverse reactions during the whole treatment period. Even when trials use the 3+3 design and dose escalation decisions are based on the first cycle, the estimation of the MTD can incorporate toxicities across all cycles in a longitudinal or time-to-event approach. The use of adaptive designs or methods such as the time-to-event continual reassessment method, which considers toxicities arising over the entire course of treatment, could provide a better estimate of tolerable MTA doses for long-term treatment. To use these methods, protocol defined DLTs will need to incorporate toxicities beyond the first one or two cycles of treatment.

The concept of tolerability is further discussed in Section 8.

Evaluation of toxicity

The general principles as discussed in Section 6.1.1 apply, but foreseeable pharmacology related adverse reactions are more diverse and should be accounted for in the planning of the studies. E.g. for immune check point inhibitors, autoimmune or immune-related reactions are foreseeable; whilst for antiangiogenic compounds vascular events, hypertension and proteinuria may be expected.

6.2.2. Phase II, single agent therapeutic exploratory studies

For the purpose of simplification, it is assumed that a dose/exposure range has been defined that shows pharmacological activity/target occupancy with or without dose limiting toxicity. If not otherwise justified, it is postulated that activities related to identification of the proper target population, as discussed above, continues in these studies.

Measures of activity

ORR, despite all its shortcomings related to patient-selection, etc., is a rather convincing measure of anti-tumour activity as for most tumours, spontaneous regression fulfilling criteria for at least partial response is a rare phenomenon. For exploratory purposes, studies without a randomised reference are therefore considered interpretable and guidance provided in the section about cytotoxic compounds is relevant. Irrespective of this, inclusion of a randomised reference arm is encouraged and might be of special interest in order to explore whether, e.g. a selected biomarker is prognostic and/or predictive (see Section 7.1.2).

Duration of response, time to progression (TTP) and progression-free survival (PFS), however, are in principle a function of underlying tumour growth rate and the activity of the anti-tumour compound.

Also, if documented progressive disease is an inclusion criterion, underlying growth rate is hard to define in most patients and historical data will be even harder to interpret. Therefore, the interpretation of TTP/PFS data without a randomised reference is problematic. However, response durations should always be reported when reporting ORR. In particular in breast cancer, clinical benefit response rate (CBR), i.e. CR, PR and absence of progression at 6 months, is a well-established measure of anti-tumour activity even though subject to the same principle problem as TTP/PFS.

Exploratory trials with time-related endpoints

There is probably no ideal yet feasible design of exploratory studies for compounds assumed to mainly elicit tumour growth control. In the following section some design alternatives are discussed, all with pros and cons, but in principle acceptable from a regulatory perspective. Irrespective of design, it is recommended that only patients with documented tumour progression are enrolled.

- A randomised, dose comparative trial, e.g. comparing the lowest dose likely to be pharmacologically active with higher dose(s), if showing a difference in TTP/PFS, will obviously provide evidence of activity, but not in absolute terms.
- Randomised withdrawal of therapy in a single arm study in patients with non-progressive disease after a defined period of time on experimental therapy. The acceptability of this design to patients and investigators, however, may constitute an obstacle and carry-over effects may be a reality for some compounds.
- In previously treated patients, a within patient comparison of TTP/PFS might provide evidence of activity. Here TTP on last prior therapy is compared with TTP/PFS on the experimental therapy. It should be noted, however, that the underlying assumption of at least similar growth rate over time cannot always be substantiated. For exploratory purposes this constitutes no major concern. It is advisable to recruit patients with secondary as well as primary resistance on prior therapy. This ensures at least to some extent, that the study population is relevant. It should also be noted that patients with early failure (primary resistance) on prior therapy may show some inversions in terms of TTP just due to fluctuations in tumour growth rate and variability related to imaging techniques.

  For certain indications a within patient comparison may be justified, also in treatment naïve patients, i.e. patients are followed without therapy until progression followed by experimental therapy until progression.
- A randomised phase II study versus a compound known to be active in the selected population (or placebo/BSC if justified) provides another alternative. In a comparison in terms of TTP/PFS it should be noted, that a purely growth inhibitory compound is "favoured" compared with a compound inducing tumour shrinkage, as progression is defined in relation to best tumour response. At the time of tumour progression, the tumour burden in patients failing a purely growth inhibitory compound will therefore be higher than in patients where tumour shrinkage was elicited.
- If no more refined techniques are applicable, TTP/PFS and CBR without an internal reference may be accepted as a measure of Phase II benefit . A systematic literature review, including methodology used, is advised in these cases.

In principle, a statistical approach similar to that for Phase II trials with ORR as outcome measure is applicable. It is harder to set up criteria for early termination, however. The number of patients should be sufficient to obtain a reasonably precise estimate of the percentage of progression-free patients at a predefined time point. The underlying assumptions as regards progression rate without therapy are more problematic and "promising activity" is harder to define.

For these studies, the use of conventional criteria for ORR and tumour progression is recommended and independent review is encouraged. It is recognised, however, that, e.g. an apparent increase in tumour size due to inflammatory oedema, "pseudoprogression", might be a first sign of activity for

certain compounds. If prior trials indicate that this is the case, it is accepted that this is accounted for in the study protocol. The use of ORR and TTP as key measures of activity should not be regarded as contradictory to the use of tumour/PD markers in parallel.

If a randomised design is considered appropriate, the use of generally accepted instrument to estimate HRQoL or symptom control may provide valuable information (see Appendix 2).

For window of opportunity studies and if sensitive measures of pharmacological activity are available, e.g. functional tumour imaging and/or biomarkers, and a target population has been identified with tumours likely to be sensitive, placebo-controlled trials with one or preferably more doses of the experimental compound might be feasible. Sensitive measures, even if not fully validated with respect to relationship to ORR, are from a regulatory perspective acceptable for exploratory purposes and allow not only for refined dose comparisons, but also early escape in case of absence of activity. It is advisable though to clearly define in the protocol criteria for progressive disease, whether a composite (e.g. biomarkers, or imaging, or symptoms) is used or not.

6.3. Monoclonal antibodies (MoAb) and immune-modulating compounds

This section is primarily meant to provide guidance as regards exploratory studies, but also on some aspects of relevance for confirmatory studies.

6.3.1. Monoclonal antibodies

Monoclonal antibodies may affect tumour cells directly, e.g. through antibody-dependent cell-mediated cytotoxicity (ADCC) and/or blocking of growth factor/anti-apoptotic receptor signalling, or indirectly through the targeting of growth factors for the tumour or tumour supportive structures, or by blocking T cell inhibitory signals (e.g. anti-CTLA4, anti-PD-1, and anti-PD-L1).

In vitro non-clinical studies should be performed to elucidate the prime activity of the MoAb. These studies may include relevant assays on:

1. Binding to target antigen(s): tumour cells or plasma should be screened for (over)-expression of the target and the relationship between target expression and activity should be investigated.
2. Unwanted targets. Tumour specificity may not be attainable, but it is possible to screen for "unwanted" targets in vitro, facilitating the safety assessment.
3. Fab-associated functions (e.g. neutralization of a soluble ligand, receptor activation or blockade).
4. Fc-associated functions (e.g. antibody-dependent cell-mediated cytotoxicity, ADCC; complement- dependent cytotoxicity, CDC; complement activation).

Target-mediated disposition may be seen with MoAbs. Adequate characterization of this form of non-dose proportional PK behaviour may not be possible until late phase studies, when patients with tumours having widely variable amounts of target are studied. Therefore, continued evaluation of MoAb PK during the clinical development program, which often involves different tumour types and stages of disease is encouraged."

Clearance of MoAbs is typically influenced by the neonatal FC receptor (FcRn) immunoglobulin G(IgG)re-cycling, immunogenicity (Anti-Drug-Antibodies (ADA)) and may also be impacted by patient health status factors (e.g. albumin, soluble receptors/ligands, disease type and severity, tumour burden, etc.). Knowledge of these factors may contribute to understanding the nature of MoAb exposure and response. The experience as regards immunogenicity of MoAbs in other fields of clinical medicine should be taken into account with respect to choice of assays, markers for loss of activity and possible safety problems.

6.3.2. Immune-modulating compounds including tumour vaccines

Immune therapies including therapeutic cancer vaccines are aimed to induce specific anti-tumour immunity toward existing malignant disease. Such immune therapies are normally aimed to induce adaptive T and B cell as well as innate immune responses in cancer patients. The nature of the drug substances used is highly variable, including synthetic peptides, recombinant proteins, virus-like particles, immune-modulating antibodies, gene therapy, and cell-based products. As it is difficult to break tolerance towards tumour antigens which are normally derived from self-antigens, cancer vaccines are often combined with pharmacologically active adjuvants such as cytokines or toll-like receptor agonists. One other approach to break immune tolerance is to block T cell inhibitory signals, e.g. with monoclonal antibodies. The resulting T-cell activation and proliferation leads to wanted and unwanted immune stimulatory effects: the desired anti-tumour effect as well as the appearance of immune related toxicities like colitis and endocrine insufficiency.

Non-clinical in vitro and in vivo proof-of-concept studies should be presented to justify the planned starting dose and schedule in phase I studies. Furthermore, and on a case-by-case basis, the rationale for the starting dose may be supported by using the 'Minimal Anticipated Biological Effect Level' (MABEL) approach, and by non-clinical and clinical data from related compounds (EMEA/CHMP/SWP/28367/07).

It is acknowledged that for products relying on human-specific antigens which need to be presented on human MHC molecules, predictive animal models are often not available. Nevertheless, animal models using homologous antigens or animals being human MHC transgenic might be considered for non-clinical pharmacology and toxicology studies, if available. Information on the differential expression of the target antigen in human tumour and healthy tissues should be provided. In case that no relevant and predictive animal model is available, in vitro studies with human cells, like e.g. in vitro T-cell priming assays, might be suitable to show proof-of-concept.

The aim of early clinical trials is to determine the safety and the dose and schedule that induced a desired immune response. Dose-finding studies are generally required to establish the recommended phase II dose. Monitoring the immune response, i.e. the induction of antigen-specific T cells or the presence of a humoral response are of interest to determine appropriate dose and schedule. To achieve this goal multiple monitoring assays may be necessary, and these should be carefully explored. The analytical methods should be described in detail in the clinical trial protocol.

Tumour biopsies taken before and after treatment are expected to play a pivotal role in assessing the extent and type of immune activation in the target tissue and could serve as an early marker for possible anti-tumour activity.

The induction of tumour response in patients with high tumour burden might be a too high hurdle to overcome and may favour the inclusion of patients with minimal or low tumour burden. Examples are therapy of patients with NSCLC after complete tumour resection where cancer immunotherapy can be assessed in the adjuvant setting. Another example is patients suffering from non-resectable NSCLC who have responded to chemotherapy. The design of clinical studies using clearly experimental therapies in patients with limited and measurable disease, not heavily pretreated with cytotoxic regimens has to be carefully justified. As for other agents, evidence of anti-tumour activity is essential prior to the initiation of confirmatory studies.

Oncology patients are usually taken off treatment upon disease progression. Induction of an effective immune response and clinical response may need more time to develop (delayed effect) compared to classical cytotoxic compounds. Patients may thus experience disease progression prior to the onset of biological activities or clinical effects. Discontinuation of active cancer immunotherapy in case of slow progression may not be appropriate. In these situations, a detailed definition of "slowly progressive disease" and/or withdrawal criteria is expected in the study protocol and close monitoring of patients is required. The definition of "slowly progressive disease" should be guided by the course of disease under investigation. Revised criteria defining progression is accepted if properly justified, in confirmatory studies, however, OS is the recommended outcome measure.

Possible toxicities like induction of autoimmune reactivity (cellular and humoral) and induction of tolerance should be carefully monitored during the clinical development.

6.4. Combination therapy studies

Conventional cytotoxic compounds have for long been used in combination in order to increase the anti-tumour activity at acceptable levels of toxicity. This may be accomplished by combining compounds with at least partly non-overlapping toxicity and, perhaps, partly non-overlapping prerequisites for activity/resistance. Regulatory agencies, as well as learned societies, have accepted this approach, but it is acknowledged that it is frequently unknown whether combined use results in a better long-term outcome than consecutive use.

6.4.1. Combining conventional cytotoxic compounds

In the selection of patients with available alternative therapies, the documented activity of the individual components of the combination regimen should be taken into account.

The exploratory phase encompasses the determination of MTD and RP2D for the combination and a preliminary assessment of anti-tumour activity in terms of ORR and PFS/TTP. While the degree of anti-tumour activity for a new combination relies on assumptions, it is often possible to predict toxicity, based on the toxicities of the individual components. If relevant PK interactions can be excluded, and depending on the dose-response/toxicity profiles, dose-finding studies may be initiated at about 1/2 of the recommended mono-therapy dose for each compound. It might also be appropriate to start at the full recommended mono-therapy dose for one of the compounds and reduced dose (<50%) for the other compound. As the sequence of administration may be of importance with respect to potential PK interactions and anti-tumour activity, this has to be accounted for in the design of the studies.

There is no uniform way to balance dose intensity between components of a combination regimen to optimise benefit – risk. It is thus accepted that, e.g. priority in terms of dose intensity is given to the compound with the highest monotherapy activity.

If one of the components is regarded as an acceptable treatment regimen in monotherapy, a randomised phase II study comparing the monotherapy regimen with the combination is informative. For confirmatory studies a comparison with the best available, evidence-based reference regimen is expected.

6.4.2. Combinations involving a non-cytotoxic drug

If there are no strong biological/pharmacological arguments to the contrary, the selected chemotherapy regimen to be combined with the non-cytotoxic should normally be "best available". If the dose intensity/systemic exposure of the chemotherapy regimen is unaltered it can be assumed that all patients will receive appropriate therapy. Therefore, there is no need to restrict the eligibility of patients from this perspective.

Whenever previous non-clinical and clinical experience has suggested that PD markers, etc. might be informative with regard to anti-tumour activity, they should be part of the experimental plan. This may include investigations whether the expression of the target for the non-cytotoxic compound is affected by treatment with cytotoxic agents and if appropriate vice versa.

Given the current status with respect to predictability of add-on activity in non-clinical models, randomised phase II studies comparing the experimental regimen with the chemotherapy-alone regimen are considered essential. For these studies, it is recommended that conventional anti-tumour activity data (ORR and TTP) are supplemented with tumour markers and sensitive measures of, e.g. tumour metabolic activity as appropriate.

When add-on activity of the non-cytotoxic compound to a chemotherapy regimen has been demonstrated, the need for further randomised phase II studies when new indications are studied may

be dispensable. This, however, should be justified as the importance of target expression and inhibition thereof might differ between malignancies.

If the expression of the target for the non-cytotoxic compound may be differently affected by different chemotherapy regimens, it is advisable to study target expression during treatment with a new chemotherapy regimen prior to the conduct of add-on studies.

Research aiming at understanding the mechanisms and prerequisites for the add-on effects is encouraged, as it may allow for an improved characterisation of target populations in future studies.

It is conceivable that for some non-cytotoxic compounds, combinations are needed not only to optimise anti-tumour activity, but actually are required in order to obtain activity. For such compounds, e.g. target saturation in monotherapy and, importantly, non-clinical toxicity for the combination may be used to define suitable starting doses and schedules. Otherwise dose/schedule exploratory and therapeutic exploratory studies may proceed essentially as for a monotherapy regimen.

If supported by strong biological and/or pharmacological non-clinical and early proof-of-principle clinical data, two new compounds may be combined in a co-development program.

The following three scenarios are foreseeable:

Uni-enhancement refers to scenarios when one combination partner *B* has no or minimal anti- tumour activity per se, but enhances the anti-tumour activity of the other partner *A* (e.g. through prevention of resistance development). The contribution of *B* needs to be established by data from appropriate non-clinical models. In phase II the comparison to a reference treatment is encouraged, while Phase II monotherapy data for *B* may be considered dispensable. An appropriate phase II design would be a randomised three-arm study *AB vs. A vs. reference treatment*.

Co-enhancement is considered when both combination partners demonstrate (modest) anti-tumour activity per se and the anti-tumour activity of the combination is considerably increased. In phase II, the new combination should be compared to both combination partners as single agents at efficacious doses and preferably a reference treatment: *AB vs. A vs. B vs. reference treatment*. Depending on the phase II results one or both monotherapy arms may be dispensable in phase III.

In case the monotherapy arm of one combination partner (B) is part of phase III (*A+B vs. B vs. reference*) the same monotherapy may not need to be included in phase II (*A+B vs. A vs. reference treatment*).

Synthetic lethality refers to a scenario when both combination partners have no or minimal anti-tumour activity *per se* but exhibit potent activity as a combination. If non-clinical and clinical studies indicate "inactivity" at dosages/exposure levels considerably above that of the combination and the combination is clearly active, the contribution of both partners may be dispensable for phase 2 and phase 3 studies.

As the same targets may have a different impact in different malignancies the necessity of both combination partners may need to be shown for new indications.

Evaluation of toxicity and tolerability in dose-finding combination studies

Irrespective of class of medicinal product and if there are no informative pharmacodynamics endpoints suitable for dose optimization, dose finding essentially relies on toxicity and tolerability. The dose finding study design depends on the class of drug, as outlined above including e.g., the need for prolonged treatment and DLT/safety observation time in order to identify dose limiting but late adverse reactions of many non-cytotoxic agents.

As discussed above, the optimal dose intensity of the individual compounds being part of the regimen is rarely possible to empirically identify from an efficacy or from a safety perspective. For combinations where co-enhancement of pharmacology activities and worsening of the safety profile of the combination compare to single partner are anticipated, particular attention should be paid to the need for a dose finding combination study prior to conduct of phase II studies. Comprehensive PK/PD

assessment for potential interactions and characterisation (also mechanistically) of on- and off-target toxicities are particularly pertinent in combination studies. Apart from identifying a regimen that is tolerable, aims should include the identification of the product(s) causing the observed adverse reactions in order to guide dose reductions in relation to observed toxicity. The toxicity profile of the drugs used as monotherapy provides some guidance, but class experience, mode of action, etc. should also be considered.

7. Phase III, confirmatory trials

Confirmatory trials should be designed with the aim to establish the benefit - risk profile of the experimental medicinal product, including supportive measures, in a well-characterised target population of relevance for clinical practice.

In the general parts of this section (Section 7.2 – 7.4), the aim of therapy, curative versus long term disease control vs. palliation and not the underlying disease has been used to structure the discussion.

For some malignancies where treatment is administered without curative intent, there are alternative, in clinical practice still well-established regimens, showing major differences in anti-tumour activity. This reflects that selection of therapy in the clinic is guided by efficacy and safety. It is therefore of relevance in the planning phase to consider the expected tolerability/toxicity profile of the experimental regimen compared with the selected reference regimen. It is fully acknowledged that safety data may be rather limited prior to the conduct of the first confirmatory trial, but main toxicities should normally have been identified and this should be sufficient for a rough estimate of the expected relative toxicity of the experimental regimen compared with alternative reference regimens.

Three categories are used in this document: Reduced or similar toxicity, increased toxicity and major increase in toxicity. No precise definition is given here due to heterogeneity of the conditions. "Major increase in toxicity", however, in most cases refers to a fear that the experimental regimen might be associated with an increase in treatment related deaths, irreversible adverse events with a long-term impact on quality of life (QoL), or severe impairment to patient condition. Other issues to consider include risk for secondary tumours. This categorisation is mainly meant for guidance in the planning of confirmatory studies and in order to provide advice on regulatory expectations with respect to study outcome measures in order to enable a proper benefit – risk assessment.

7.1. Design

7.1.1. Patient population

With respect to diagnosis, criteria for initiation of treatment, eligibility, response criteria and choice of reference therapy, a justification based on scientific evidence and/or generally acknowledged and updated treatment guidelines are expected. While this is true in general, it is also expected that the exploratory studies through the judicious use of biomarkers provide guidance with respect to selection of patients in order to optimise benefit – risk, whether patient selection is in need for confirmation or not, in the planned phase III trials.

There is a general wish to reduce heterogeneity of study populations (performance status, co-morbidity, organ dysfunction, etc.) in order to increase the ability of the study to detect differences between study arms. This has to be balanced against the availability of patients for inclusion and the wish to enrol a clinically representative selection of patients. Therefore, investigators should normally be encouraged to include patients representative of those likely to be treated with the experimental compound in clinical practice. Restrictions as regards, e.g. performance status should be reflected in the SmPC. With respect to studies with a non-inferiority efficacy objective, please refer to7.6.4.

Patients are expected to be characterised by relevant tumour parameters, e.g. stage, grade, target expression, other biomarkers of importance for prognosis and/or tumour sensitivity, prior therapy (responsive/ resistant/refractory as appropriate), as well as performance status, co-morbidity, organ dysfunction, etc. Stratification based on important and well-established prognostic covariates should be

considered. In case adjusted analyses are to be undertaken for covariates other than those used for stratification, these factors should be pre-specified in the protocol or the statistical analysis plan (CPMP/EWP/2863/99).

If exploratory studies provide a basis for including/excluding certain patients based on tumour phenotype/genotype, this will be reflected in the labelling. As a corollary, if patients with tumours not expressing the target for activity are eligible, a restricted labelling may still be appropriate if it has not been demonstrated, e.g. by subgroup analyses, that target expression is irrelevant for anti-tumour activity.

If it is expected that a biomarker defining eligibility to the trial will be assessed locally or regionally in clinical practice, it is recommended that this is done also for the trial, complemented with central assessment of the biomarker to make feasible sensitivity analyses, etc.

As some of the conditions are rare, it is understood that the Sponsor might wish to define the target population using alternative criteria to those commonly employed. For example, in studies investigating the activity of a compound targeting a specific, molecularly well-defined structure assumed to be pivotal for the condition(s), it might be possible to enrol patients with formally different histological diagnosis but expressing this target.

The pivotal role of the target in different histological diagnoses, however, must be demonstrated. This should be addressed in clinical studies, but it is accepted that formal testing with adequate statistical power of such a hypothesis cannot always be done. Possible consequences with respect to selection of proper reference therapy(ies) must be considered and the study should be designed so that it is possible, based on all available evidence, including non-clinical and pharmacological data, to conclude on the benefit – risk in the different subgroups of patients for which a claim is to be made. Prior to the initiation of confirmatory studies using non-conventional criteria for eligibility, EU scientific advice should be sought.

Some possible target indications comprise very small groups of patients, so small that a marketing authorisation under "exceptional circumstances" might apply. Unless the target for activity is expressed only in these rare conditions, Sponsors are in general advised to undertake studies in these small patient groups in parallel to or when benefit – risk is established in indications allowing a more comprehensive evaluation, especially with respect to safety.

7.1.2. Reference therapy

The choice of reference regimen should be justified and normally this regimen should be selected from best available, evidence-based therapeutic options. In this context, "best available, evidence-based" should be read as a widely used, but not necessarily licensed regimen with a favourable benefit-risk convincingly documented through randomised trials and considered at least as good from a benefit/risk perspective as alternative, treatment options.

It is acknowledged that there are different, region-preferred standards. For superiority studies (test vs. reference) this should normally not constitute a problem as long as the reference is evidence-based as defined above. For add-on studies (reference + test vs. reference), it might also be possible to use a few, region-preferred references. Here a convincing clinical/pharmacological justification is needed, and EU scientific advice is recommended. Whenever more than one reference regimen is used, stratification is recommended.

If the aim is to demonstrate non-inferior efficacy, the selected reference regimen must enable a proper definition of the non-inferiority margin. In most cases, this would require that randomized well-controlled studies have shown the superiority of the selected reference vs. control. Please also refer to Section 7.6.4.

Amongst best available references, regimens with similar cycle lengths should be prioritised as it facilitates the identical scheduling of tumour assessments. If the objective is not to improve tolerability and toxicity, a regimen with similar expected toxicity to the experimental regimen is also preferred. This might also make the conduct of the study under double-blind conditions possible, a design

recommended whenever adverse reactions do not make attempts to blind the study futile. In add-on studies (to an active reference or BSC), placebo is also recommended whenever meaningful.

In some cases, there is no well documented reference regimen, even though patients in clinical practice are treated with certain regimens. Even though BSC is acceptable in these cases, an active comparator, documented e.g. in terms of response rate, is often preferable. If a single reference regimen cannot be defined, investigator's best choice is an option. In these cases, reference regimens with low toxicity are favoured and superiority in terms of patient relevant endpoints should be demonstrated.

The absence of evidence-based therapies often refers to patients who have failed several lines of therapy. In this situation, it might be more informative and also easier to obtain the data needed for marketing authorisation based on a properly conducted randomised study in less advanced patients, supported by "salvage" single arm studies, compared with conducting a last line, randomised BSC/investigator's best choice comparative study.

Single agent and combination therapies

Whether the experimental agent is used as a single agent or in combination, the experimental regimen should be compared with the "best available" comparator again referring to benefit/risk, not only to efficacy.

If the experimental agent (A) is added to an established regimen (B), superiority of AB vs. B should be demonstrated, and benefit-risk should be shown to be favourable. A discussion is expected based on available data as regards dose intensity of B and benefit risk. Traditionally, this type of studies does not include an A alone third arm, but this should be justified based on available exploratory study data.

In case of substitution studies, i.e. studies where a component (C) of an established regimen (BC) is replaced with an experimental agent (A) and if non-inferiority (BC vs. BA) is the aim, the contribution of C to the activity of BC has to be well defined (CPMP/EWP/2158/99).

Uncommonly, an entirely new combination AB is tested against a reference regimen. In these cases, solid non-clinical and clinical phase I/II data should support the need for both components in the experimental regimen.

7.1.3. Cross-over

In order to enable a qualified benefit – risk assessment, cross-over at time of progression should be undertaken only when detrimental effects on OS have been excluded (see Appendix 1).

7.1.4. Randomisation and blinding

Randomisation and stratification should adhere to the general principles laid down in current guidelines (CPMP/ICH/363/96). In many cases, a double-blind design is no option due to obvious differences in toxicity between study regimens or due to safety concerns. If the study has to be conducted open label, this has implications with respect to choice of study endpoints, independent review, conduct of sensitivity analyses and other measures to be undertaken to limit potential bias related to the open-label nature of the trial.

7.1.5. Endpoints

Confirmatory trials should demonstrate that the investigational product provides clinical benefit. There should thus be sufficient evidence available demonstrating that the chosen primary endpoint can provide a valid and reliable measure of clinical benefit in the patient population described by the inclusion criteria. In the following, superiority trials aiming to establish efficacy are the focus of the discussion.

There are a number of clinical endpoints, which are considered as adequate primary endpoints in confirmatory clinical trials to measure clinical benefit. These typically include OS, PFS, EFS, and DFS. Selected patient-reported outcomes (PROs), such as symptom control, could also constitute clinically relevant and valid primary endpoints, provided high data quality is ensured. In some situations, other primary endpoints have also been considered as appropriate, such as enabling further treatments known to be beneficial (e.g., stem cell transplantation) or avoiding treatments considered to be associated with high morbidity or mortality (e.g., invasive surgery).

Generally, convincingly demonstrated favourable effects on survival duration are, from both a clinical and methodological perspective, the most persuasive outcome of a clinical trial.

An effect on prolonging PFS of sufficient magnitude, and provided a detriment on other important endpoints can be excluded, is considered in itself a clinically relevant effect because documented progression of the disease is generally assumed to be associated with subsequent onset or worsening of symptoms, worsening of quality of life, and the need for subsequent treatments generally associated with lower efficacy and worse toxicity. If these assumptions do not hold (e.g., if there are equally efficacious and safe "rescue" treatments available in subsequent lines) then an effect on PFS may be considered less clinically important and it may be difficult to establish a positive benefit-risk balance based on this endpoint (see Appendix 1: Methodological considerations for using PFS or DFS in confirmatory trials).

If PFS or DFS is the selected primary endpoint, OS should be reported as a secondary and vice versa. In situations where there is a large effect on PFS, (as primary objective), or where there is an expected long survival after progression, and/or a clearly favourable safety profile, precise estimates of OS may not be needed for approval, but no signs of a detrimental effect on OS should be present. Furthermore, regardless of the chosen primary clinical endpoint, any detriment or uncertainty in other important clinical endpoints, including safety, would generally be considered to impact negatively on the benefit-risk assessment.

When OS is reported as primary endpoint, consistency is expected as regards effects on PFS. If foreseen not to be the case, e.g. in case of certain immune modulating therapies, this should be made clear already in the study protocol.

For some conditions, events of progression will be observed at a slow rate making frequent assessments of events of progression a burden to the patients. Event rate at a pre-specified and justified fixed point in time might be used as primary outcome measure in these cases. When event rate at a single point in time is selected for the primary analysis, it is in most cases recommended that all patients should have been on study for that period of time. PFS in a time to event analysis, and as assessed by the investigator should be reported as a secondary endpoint when a fixed time-point assessment is used as primary outcome measure.

For further methodological guidance as regards PFS, please refer to Appendix 1.

The tumour's drug resistance profile is expected to be affected by therapy. This might be of relevance for the activity of next-line therapies, which is most obvious if maintenance/prolonged therapy is compared with no treatment or placebo, but also in cases with a substantially increased number of "induction" cycles compared with the current standard of care. The consequences of progression on maintenance therapy might thus differ from progression off therapy. If possible, main studies should therefore be designed with the aim to document the effect of the treatment on duration of overall survival. If non-feasible, endpoints such as PFS on next-line therapy (PFS2) should be determined (see Appendix 1). This should ideally be done within the study so that agreed next line therapy(ies) is used after progression in the different treatment groups. In order to capture possible negative effects on next-line therapy and to outbalance tolerability and toxicity concerns related to therapy, it is expected that time from randomisation to PFS2 in the experimental arm show no detrimental effect compared to the control arm. As methodological issues are foreseeable, EU scientific advice should be considered.

If the experimental compound used for maintenance therapy can be used as single agent also at time of recurrence, it is recommended that early treatment, i.e. maintenance, is compared with deferred therapy, i.e. treatment at time of progression.

It is accepted that it may not be feasible to define next-line therapy within the study protocol and to follow patients with scheduled assessments until PFS2. Time on next-line therapy might in these cases be used as a proxy for PFS2. The likely increased variability in the assessment of "PFS2" will be taken into account in the comparison PFS2control vs. PFS2exp

In general, regardless of the primary endpoint, it is recommended that reasons for selecting a certain next line therapy, and time on next-line therapy, are collected in the CRFs and presented.

In patients with tumour-related symptoms at baseline, symptom control, if related to anti-tumour effects, is a valid measure of therapeutic activity and may serve as primary endpoint in late line therapy studies. In certain cases, symptomatic progression-free survival may also be an adequate primary measure of patient benefit.

HRQoL/PROs can provide important patient perspectives on the disease and the treatment received. Clinical studies to support regulatory submissions are encouraged to include relevant PRO measures, as secondary or exploratory outcomes or as primary outcomes when justified, using carefully validated tools. Careful planning and analysis of how the inclusion of PRO measures is likely to make a potential difference to the interpretation of the study results is key (see Appendix 2: The use of PRO measures in oncology).

There are also examples where tumour response-related activities, e.g. limb-saving surgery may be reasonable primary measures of patient benefit. Analyses of location- or cause-specific events, however, should in general be avoided as the focus may be drawn away from the main objective, namely the overall success of the treatment strategy in question.

Irrespective of the choice of primary endpoint, ORR, DoR and rate of tumour stabilisation for, e.g. 3 or 6 months should be reported. Overall consistency in outcomes is expected across endpoints, unless justified, e.g. in terms of mechanism of action and tumour biology.

Scientific advice is recommended in cases where deviations from the guideline are planned. See also Appendix 4 (condition specific guidance) on the pathological complete response as an endpoint in neoadjuvant breast cancer studies and use of minimal residual disease as an endpoint in chronic lymphocytic leukaemia studies, as well as specific guidance for NSCLC, CML, myelodysplastic syndromes, and haematopoietic stem cell transplantation.

7.2. Treatment administered with curative intent

The ultimate aim of developing new therapies, e.g., in patients with high grade lymphoma, germ cell tumours or in the adjuvant setting, is to improve cure rate and survival or to relevantly decrease toxicity without loss of efficacy. Nevertheless, in some cases and due to the complexity of administered therapies, e.g. in AML, the impact of a relevantly active experimental compound on these endpoints may be hard to demonstrate.

It is foreseen that the experimental compound rarely will be used as single agent therapy, but will be used as add-on to an established, perhaps modified regimen, or as substitution for a compound being part of the established regimen. In this context, maintenance therapy may be regarded as add-on therapy if maintenance therapy is considered non-established.

In the treatment of acute leukaemia, lack of achievement of CR, relapse and death without relapse are counted as events in an EFS analysis. Those patients who did not reach CR during the pre-specified induction phase will be considered as having an event at time 0.

In case EFS is found to be a justified primary endpoint, it is of importance that study data are analysed only when sufficiently mature, i.e. when it is foreseen that the EFS plateau is stable or when additional disease recurrence is rare.

In patients with high grade lymphoma or solid tumours, PFS may be used as outcome measure. Not achieving at least PR after a defined period/number of cycles may be regarded as treatment failure in

some protocols and only those achieving at least PR continue on therapy. In the primary analysis it is recommended that patients not reaching PR are followed off or on next-line therapy until an event of progression or death is reached.

When improved cure rate is the objective of therapy, it is advised that disease-free survival at a pre-specified time point is used as outcome measure (see above with respect to timing).

7.2.1. Reduced or similar toxicity expected

In most cases, a substitution design is foreseen, meaning that A in an established regimen (AB) is replaced with the experimental agent X (XB). From a regulatory perspective, a non-inferiority design is acceptable and, in most cases, EFS or PFS, as appropriate, are acceptable primary endpoints.

In cases where induction is followed by consolidation and/or maintenance therapy, confounding effects of therapies administered after the end of experimental therapy may make endpoints other than PFS or EFS more appropriate. This means that CR (and CR + PR, if specifically justified) after end of experimental therapy could be an acceptable primary endpoint when further therapy is scheduled. In these cases, the possible influence of the experimental compound on the activity of consolidation therapy should always be addressed and outcomes with respect to CR should be supported by EFS or PFS data.

It is recommended that CR is defined according to established clinical criteria, but supportive evidence in terms of Minimal Residual Disease (MRD) as defined, e.g. by molecular criteria should be sought when applicable. As for other biomarkers, intra- and inter- laboratory variability should be minimised through standardisation.

7.2.2. Increased toxicity expected

Substitution or add-on designs may apply. In most cases, superiority in terms of EFS, PFS, or OS as appropriate, should be demonstrated and the benefit in terms of prolonged time to event should be sufficiently large to balance increased toxicity.

A major increase in CR after induction therapy associated with trends in PFS or EFS, and survival, however, might be sufficient if scheduled treatments administered after the end of the experimental therapy are likely to confound overall outcome. This is of special relevance if the target population is small.

7.2.3. Major increase in toxicity expected

The aim should be to demonstrate increased cure rate or improved OS. In some cases, such as in small study populations, a major increase in EFS or PFS, as appropriate and supportive data compatible with a favourable trend on survival might be sufficient.

7.3. Treatment administered with the intent to achieve long-term disease control

Typical conditions include early lines of therapy in advanced breast cancer, colorectal cancer, low-grade lymphomas and the chronic leukaemias for which established reference therapies are available and next-line treatment options are likely to be meaningfully efficacious.

7.3.1. Reduced or similar toxicity expected

Substitution or single agent studies are foreseen. From a regulatory perspective, a non-inferiority design is acceptable and PFS is considered an appropriate primary endpoint. In case of relevantly reduced toxicity, mature survival data may be submitted post licensure if justified by study data.

7.3.2. Increased toxicity expected

The aim should be to demonstrate superiority at least in terms of PFS.

Survival data should be made available at the time of submission. It is acknowledged that mature survival data cannot be expected in all cases, though a justification explaining why this is the case should be provided. Post approval follow-up with respect to survival is expected in these cases. If absence of an increase in treatment-related mortality is not established with reasonable certainty, mature survival data should be available for the assessment of benefit – risk prior to licensure.

It is acknowledged that alternative endpoints may be more appropriate in certain situations, e.g. when maintenance therapy is investigated in areas where this has not established (Endpoints, 7.1.5). The aim may also be to enable a long treatment-free interval after intense induction therapy.

7.3.3. Major increase in toxicity expected

The principal objective should be to demonstrate improved survival.

In individual cases this might be non-achievable due to expected good prognosis with respect to survival and availability of several active next-line regimens, including experimental therapies, at the time of disease progression and a small target population. If PFS is the selected primary endpoint for the study, this requires a thorough justification. A careful discussion at the planning stage is also needed for the assessment of possibly therapy-related fatalities. Even though only a major benefit in terms of PFS prolongation would be acceptable, whenever possible the number of patients included should be sufficient to obtain an estimate on overall survival where a trend in a favourable direction is expected.

7.4. Treatments administrated in settings with lack of established regimens

This mainly refers to last line settings where the prognosis for survival is poor and where it might be problematic to identify sufficiently documented reference therapies. In other cases, patients are considered not suitable for intensive, potentially curative therapy as defined by clear and as far as possible unambiguous criteria.

In cases where there is no established reference therapy, investigator's best choice or BSC with or without placebo are acceptable.

In a study conducted with BSC as reference therapy, the objective of demonstrating prolonged OS and/or globally improved symptom control or HRQoL, is particularly important. The latter requires that all efforts are undertaken to reduce possible bias (Appendix 2).

7.5. Special considerations

7.5.1. Haematopoietic stem cell transplantation, methodological considerations

If allogeneic haematopoietic stem cell transplantation (HSCT) is a foreseeable treatment option, it is of importance to define how transplantation should be handled in the analysis plan. It is fully acknowledged that criteria for HSCT (e.g. patient eligibility, HLA matching, conditioning regimen, graft versus host disease prevention, etc.) vary between institutions and regions. Nevertheless, these criteria should be defined as far as possible in the protocol and reasons for performing or not performing HSCT should be captured by the CRF.

Even though transplant related mortality is an issue and long-term benefit requires prolonged follow-up, it is normally expected that patients undergoing HSCT are followed for OS and EFS as randomised. Patients may be censored at time of conditioning for HSCT as a sensitivity analysis.

Autologous stem cell transplantation constitutes less of a concern from an assessment perspective and may be viewed as intensified consolidation therapy where the consequences on short-term mortality and possible long-term benefit are less pronounced than after HSCT. Nevertheless, heterogeneity in the conduct of autologous transplantation should be avoided as far as possible, and censoring should not be undertaken.

With respect to drug development specifically in relation to HSCT, please refer to Appendix 4.

7.5.2. (Neo)adjuvant therapy

In the adjuvant setting, the ultimate aim is to increase cure rate. While effects on DFS are considered a benefit to the individual patient, regardless if cure is achieved or not, it is of importance to consider in the planning of the study whether it is at all possible to demonstrate a favourable effect on cure rate, i.e. in analyses conducted when recurrence rates have reached an apparent plateau.

As the use of adjuvant therapy may limit therapeutic options at time of recurrence, OS data should be reported. For established areas of adjuvant therapy, e.g. breast and colorectal cancer, and if benefit-risk is considered favourable for the experimental regimen based on DFS and available safety and survival data, including PFS on next-line therapy following recurrence of the disease, mature survival data may be reported post-licensing. In some cases, and due to major toxicity concerns, favourable effects on OS have to be demonstrated.

The objectives of neoadjuvant therapy may include improved overall outcome (OS, DFS/PFS), enabling surgery and organ preservation (e.g. more conservative surgery). If organ preservation is the main objective, at least non-inferior DFS/PFS should be documented. As for adjuvant therapy, a defined number of cycles is frequently administered. Pending on the objectives of the study it is accepted that treatment is withdrawn if tumour shrinkage is not observed after a defined treatment period.

When pathological CR at time of surgery is reported as secondary endpoint, patients withdrawn should be considered as non-responders.

7.5.3. Drug resistance modifiers, chemoprotective agents and radio/chemo sensitizers

In principle, the design of confirmatory studies for experimental drug resistance modifying agents and radio/chemo sensitizers (A) is straight forward; AB should be demonstrated to be more active than an established regimen (B) in terms of anti-tumour activity and the benefit – risk for the combination should be shown to be favourable. If there are PK interactions, or dynamic interactions not related to anti-tumour activity, dose adjustments of B in the combination arm might be needed in order to make the comparison AB vs. B at similar overall toxicity. If the full effects of the PK interaction is captured by changes in the plasma levels of B (e.g. no changes in distribution), however, dose adjustments of B in order to compare AB vs. B at similar exposure of B is preferred.

For a chemoprotective agent, it has to be shown that normal tissues are more protected from toxicity than tumour tissue. For most cytotoxic compounds, it is, however, easier to detect dose-related differences in toxicity than in efficacy. This means that in many cases very large studies are needed with tight confidence intervals around measures of anti-tumour activity in order to prove that normal tissue protection is achieved without loss of anti-tumour activity. Co-primary endpoints are thus needed, testing the hypotheses of improved safety and non-inferior anti-tumour activity. In some cases, it might actually be easier to convincingly demonstrate differential tissue protection by increasing the dose of the cytotoxic compound in the experimental arm aiming to show enhanced anti-tumour activity without increased toxicity.

However, if it can be shown conclusively that there is no PK interaction and that the chemoprotective compound cannot interact with the tumour, e.g. by absence of target in tumour cells, it might be acceptable only to show reduced toxicity without formal non-inferiority testing of tumour protection.

7.5.4. Tumour prevention

Regulatory experience is limited, but conceptually the situation is rather similar to the adjuvant setting. Thus individuals at risk should be defined so that the observed risk reduction in tumour incidence outweighs the side effects of therapy. As tumour prevention may select for tumours with altered biological behaviour, comparative data on tumour pheno/genotype are expected and data on tumour response to therapy or OS may be needed. In the planning of these studies, regulatory scientific advice is recommended.

7.6. Methodological considerations

Frequently, only one single study is foreseen for a specific indication. Licensing based on one pivotal study, however, requires demonstration of efficacy at levels beyond standard criteria for statistical significance (CPMP/EWP/2330/99). This is of special relevance in non-inferiority trials, in trials with PFS as primary endpoint and in a comparison with BSC/investigator's best choice. It is acknowledged that supportive evidence from confirmatory studies conducted in other indications should be taken into account in the assessment. The supportive value of these studies might vary and a discussion is expected as regards the relevance of these findings in relation to the application for the new indication.

7.6.1. Adaptive design

If a phase II/III study is designed only to address a single and non-complex question in phase II of the trial, such as proper dose for the confirmatory stage, adaptive design might increase the efficiency of drug development.

Whenever more complex issues are to be addressed, e.g. involving defining the proper target population, or multiple issues, e.g. sample size re-estimation and cut-offs for biomarker positive tumour samples, etc. it is questioned whether adaptive design approaches are advantageous and scientific advice should be considered. The need for independent supportive efficacy/safety studies as part of the application for marketing authorisation should also be considered (see Points to consider on application with 1. Meta-analyses; 2. One pivotal study CPMP/EWP/2330/99).

7.6.2. Interim analyses

Interim analyses are frequently undertaken in Phase III trials, but early stopping whether for futility or superiority is a sensitive issue. Early stopping for superiority requires that the treatment effect in patients with rapidly progressing tumours ("early events") is similar to that in less aggressive tumours ("late events") in the absence of data actually demonstrating that this is the case.

In every case, the expression of maturity must clearly refer to the number of observed events compared to the total number of events expected in the included population, and not only with reference to the proposed final analysis or to the timing of the interim analysis compared to the duration of the trial.

If a clear majority of the total number of expected events in the long term has been observed and a difference has been documented, this is normally accepted as an indicator that the study is reasonably mature and that the study results will remain stable over prolonged follow-up. The interpretation of interim analyses conducted on a less mature data set is problematic.

In cases where the treatment effect has been underestimated in the planning of the study, this may create a dilemma if statistically convincing effects in terms of overall survival have been demonstrated before a representative and mature dataset is available. Other monitoring committee decisions might be investigated in this instance such as restricting the continuation of the trial to the under-represented subsets to which the observed effect cannot be extrapolated. Analyses according to stratification factors of major importance for prognosis might provide insights as well as similar analyses with respect to PFS.

Proposals for early interruption for efficacy should be pre-specified and receive support from evidence demonstrating that prolonging the study would not significantly change the perception of benefit. This is in addition only justified if the benefit established from the early analysis is so important that the control arm is no longer acceptable for all patients matching inclusion criteria.

In general, interim analyses based on PFS data are not encouraged (Appendix 1).

7.6.3. Time to event analyses and assessment of response and progression

For studies with PFS/DFS as primary endpoint, symmetry with respect to imaging and study visits is pivotal and adherence to protocol-defined schedules is essential and deviations should be reported (Appendix 1).

Differences in mode of action between the experimental and reference therapy might generate problems in relation to measurements of tumour burden and anti-tumour activity, one example being early tumour swelling as discussed previously. Whenever such problems are foreseen, which may require deviation from standard approaches (RECIST, WHO), it is recommended that agreement is reached with regulatory agencies prior to the initiation of pivotal trials. Similarly, if tumour assessment techniques cannot be used that allow for independent adjudication, it is advisable to discuss available alternatives with regulatory agencies.

Pseudo-response should always be considered a possibility when tumour related oedema is an issue such as in high grade gliomas. Updated response and progression criteria taking this into account should be applied when available. If such criteria have not yet been established, scientific advice is recommended in order to discuss alternative ways forward.

7.6.4. Non-inferiority studies

Guidance about the design, conduct, and analysis of non-inferiority studies is given in other regulatory guidance documents (Choice of a Non-Inferiority Margin: CPMP/EWP/2158/99), but some topics deserve particular attention in the oncology setting. For a PFS endpoint, which can be considered a composite endpoint, the discussion of a non-inferiority margin should consider the effect of the reference treatment overall but inference should also include a discussion on each type of events (death, new metastases, progression of target lesions, clinical progression) including description of the effect of the reference regimen on each component when available. If differences in the profiles of progressive disease might be expected, this should be accounted for in the planning stage with a suitably conservative margin and appropriate sample size to obtain the required number of events for reliable inference.

Given the importance of study sensitivity (i.e. the ability of a trial to detect differences) for the assessment of non-inferiority trials, where similar activity is assumed for test and reference, it is of importance to plan in advance for a subgroup analysis, e.g. excluding patients with poor prognostic factors at baseline such as poor PS, co-morbidities, etc. as in these patients it might be harder to detect a difference in activity between treatment regimens, if there were one. Similarly, a per protocol analysis set should be defined so that protocol violations, compliance problems, etc. do not reduce the possibility to detect a difference. These analyses are expected to be undertaken with the aim to show consistency of the results between the study populations.

7.6.5. Analyses based on a grouping of patients on an outcome of treatment

Comparisons of time-to-event variables (like OS, or PFS) by grouping patients on a post-randomisation outcome of treatment are problematic. Since outcomes like tumour response, dose intensity, toxicity, or compliance represent an interaction between therapy, patient and tumour the contribution of therapy cannot be disentangled. Nevertheless, certain unexpected outcomes such as clearly improved survival despite dose-reduction due to toxicity, or absence of prolonged survival in responding patients

might be informative. A search for unexpected findings constitutes a rationale for conducting these exploratory analyses.

Response duration comparing groups of patients on different therapies may be regarded as informative. Data should be reported with confidence intervals for the individual study arms, but significance testing comparing duration of response between study arms should not be undertaken as the comparison refers to groups that are not fully randomised. "Time in response" where patients without response are assigned a duration of zero enables a statistical comparison between study groups.

7.6.6. Use of external control

The use of external control (including historical control) is discussed in ICH Topic E10: Choice of control in clinical trials (CHMP/ICH/364/96) and it is concluded that "the inability to control bias restricts use of the external control design to situations where the treatment effect is dramatic and the usual course of the disease highly predictable".
In these cases, prospective confirmation in randomized, reference-controlled studies is not only unacceptable to investigators, patients and ethics committees, but also unnecessary.

7.7. Special populations

7.7.1. Elderly and frail patients

Whenever elderly patients are expected to be treated with the new medicinal product in clinical practice, the clinical studies program should enrol a sufficiently large number of elderly patients, including those with co-morbidities, to enable a benefit – risk assessment. It is acknowledged that for some products, the safety of the drug needs to be established in otherwise healthy patients prior to enrolment of less fit elderly in confirmatory studies, but a justification is expected in these cases. Of note, eligibility criteria per se is frequently not the hurdle, in order to accomplish a fair representation of elderly, investigators need specific encouragement and support to enrol these patients.

It is expected that all reasonable efforts are undertaken to provide informative data in the MAA, however, if benefit – risk cannot be assessed with reasonable certainty in elderly patients or those with prevalent co-morbidities in the target population, this should be reflected in the labelling and post approval studies may need to be undertaken. In this context it is noticed that also well-planned cohort studies may provide valuable information.

Data from elderly patients should be available for pharmacokinetic analyses, e.g. as part of population pharmacokinetic analyses. Description of the safety profile should include aspects of severity of the adverse events profile and consequences, e.g. dose reduction, dose delay or initiation of concomitant treatment. An evaluation of the consistency of treatment effects and safety profile in elderly population, including age groups as appropriate, with the younger population(s) is expected.

Some compounds may be specifically suitable for the treatment of elderly, e.g. due to PK properties such as low sensitivity to impaired organ function. In these cases, dedicated studies in the elderly are encouraged. It is acknowledged that it may be hard to identify appropriate reference therapies in some of these cases and that other outcome measures than PFS/OS might become more relevant. In these cases, it is advisable to seek regulatory agreement on the development program.

Frail patients, whether elderly or not, with clearly impaired performance status (PS) constitute a vulnerable group of patients rarely included in conventional studies. Clinical studies in this group of patients are supported from a regulatory perspective.

7.7.2. Children

Paediatric cancers are all rare or very rare entities. The available guidance in the main text above and in the relevant appendices, e.g. on the use of single arm trials, biomarkers, innovative trial designs, PROs, etc. applies also to the paediatric setting.

Notably, the EU Paediatric Regulation (Regulation (EC) No 1901/2006) requires consideration of paediatric development early during the development process to ensure timely access for neonates, infants, children and adolescents to innovative treatment.

Further guidance on specific aspects of paediatric medicinal product development is available in the dedicated guidelines listed in Section 3.

7.7.3. Sex

For some tumours and/or therapies, a difference in anti-tumour activity related to sex has been reported. Where a priori it is likely that there may be a treatment by gender interaction, this should be considered in the design of the study. Otherwise it is expected that the proportion of females and males reflects the prevalence of the disease and that the sponsor provides exploratory subgroup analyses (efficacy and safety) by sex.

7.7.4. Patients with impaired organ function

Please refer to Section 4, Pharmacokinetics.

8. Specific designs for special situations

8.1. Studies in small study populations, very rare cancers

This section presents considerations regarding the investigating products targeting very rare cancers or narrow indications. Very rare cancer in this context relates to cases where, due to cancer phenotype or restrictions related to target expression, it is simply not possible to recruit a sufficiently large number of patients to conduct reasonably powered randomised studies.

A randomised clinical study is expected whenever feasible. There are several factors which could be tuned to increase the feasibility of demonstrating both a statistically significant and clinically relevant treatment effect despite a limited number of patients. For example, the choice of the primary endpoint, the length of the follow-up period, and/or the selected population, could be optimized in order to increase the statistical power of the study. Even if not sufficiently powered, randomised studies are usually preferred as they might allow obtaining an unbiased treatment effect, as well as comparative safety data.

There are situations where the feasibility of conducting a RCT is very limited despite adjustments in the study design. For those situations, alternative designs may need to be considered (see Clinical trials in small populations, CHMP/EWP/83561/2005; and Choice of control group and related issues in clinical trials, CPMP/ICH/364/96).

Resorting to non-randomized trials should be duly justified (e.g. predictable course of the disease in combination with a large treatment effect on endpoints such as ORR and duration of response reasonably likely to translate in true clinical benefit, and acceptable toxicity). Long-term efficacy and safety should always be collected unless otherwise justified.

Important uncertainties on the effects, or on the lack of important detriment, on clinical endpoints considered to directly reflect clinical benefit (e.g. OS, PFS) would typically have to be addressed on the basis of indirect comparisons and further investigated in the post-authorisation setting, making this type of evidence challenging, even in the absence of available treatment options.

Establishing efficacy and a positive benefit-risk based on non-randomized studies might be particularly challenging if there are available treatments with known effects in terms of important clinical endpoints like OS, and/or less uncertainties.

Information obtained from other sources such as real-world data or computational modelling, could complement the evidence found in the study.

Optimization of the randomization process might help to improve the attractiveness of a randomised study. For example, randomization with unequal allocation ratio or other more advanced randomization methods allocating more patients to promising study arms could be considered.

In other cases, a within-patient analysis might be an alternative where TTP on the last prior therapy is compared with PFS on the experimental therapy. This type of control suffers of similar weaknesses as historical comparisons. To minimise potential biases, it is important that the conditions under which the prior and experimental therapies were given are overall comparable. Any uncertainties related to less stringent assessment intervals and progression adjudication in the prior treatment line have to be adequately addressed. The analysis will be more convincing if prior therapy is chosen among clinically appropriate options and if progression on both prior and experimental therapy is independently adjudicated, e.g., using blinded independent central review.

Demonstrating superiority should normally be the main goal. When the objective of the study is to demonstrate non-inferiority, additional aspects have to be taken into consideration. The non-inferiority setting is generally very challenging in studies with limited numbers of patients.

In situations where a single-arm trial or an underpowered randomized controlled study is justified, contextualisation of the results is a key issue. In some cases, when the response is dramatic, occurs rapidly following treatment, and is unlikely to have occurred spontaneously (e.g., measurable tumour shrinkage), assessment may be based on general knowledge. However, in less evident cases, specific historical experience should be sought. Designers and analysts of such trials need to be aware of the limitations of studies using indirect comparisons and should be prepared to justify their use (see ICH E10, choice of control group in clinical trials).

Data could be collected from previous clinical trials, meta analyses, registry data bases or other sources, provided they are of sufficiently high quality. Preferably, patient-level data is expected as the basis for historical controls. Furthermore, it is required that these patients were treated with the current standard of care so that they form a relevant comparator for the new therapy. The historical controls are expected to be comparable also in terms of important demographic and prognostic baseline variables. Ideally, the rationale for the choice of the data sources containing the historical controls should be discussed along with their exhaustiveness. The criteria to filter historical patients out of these data sources to match the patients of the experimental arm should also be given, as well as the statistical methods used to calculate the treatment effect and to adjust for potential imbalances between the treatment group and historical controls. It is imperative that these steps are pre-specified prospectively in the protocol to avoid any convenient selection of historical controls once the endpoint has been observed in the experimental arm.

As there is no general solution to the problem of how to minimise uncertainties about treatment effects in small populations before and after marketing authorisation, scientific advice is strongly recommended.

8.2. Basket and Umbrella trials

Introduction

Alternative clinical trial designs may sometimes be warranted in situations when standard evidence-generating strategies are not feasible. The rationale for the proposed study design must be appropriately justified.

Two major types of these strategies are delineated below. As the regulatory experience is currently limited to early phase trials and these situations are often complex, it is strongly recommended to apply for scientific advice when such designs are considered for the generation of pivotal data for marketing authorisation.

The general aim for these new designs is to investigate more than one disease entity and/or more than one drug under the same trial protocol. These study designs may be generally referred to as master protocols, with further specification of the design by terms such as basket and umbrella trials.

Basket trials aim to investigate the efficacy and safety of one drug or combination of drugs in a study population that is comprised of a variety of malignant diseases and defined by the presence of a (presumed) response-predictive biomarker. The included subgroups/subpopulations may be defined by different conventional histology- or anatomy-based tumour types and are generally referred to as "baskets". Basket trials are used especially when the prevalence of the putative biomarker within a given histology is low making it difficult to enrol adequate number of patients in a conventional histology-based trial.

Umbrella trials aim to explore the activity of several drugs or their combinations in parallel within one disease entity (often anatomy or histology-based). When designed in a perpetual manner, with agents allowed to enter or leave the study according to an algorithm, these umbrella-type studies are often referred to as platform trials.

General methodology

Notwithstanding the increased logistic and statistical complexity and challenges of these designs, the general methodological and statistical considerations for exploratory and confirmatory trials described in this and other relevant guidelines are applicable (e.g. control of type I error, adaptive features, biomarker validation, dose-finding, recommendations specific to cytotoxic and non-cytotoxic compounds and their combination etc.).

For basket as well as umbrella trials, potential heterogeneity in the sub-populations may be an issue and should be prospectively addressed. Heterogeneity may for example arise from differences in natural history (e.g. spontaneous prognosis) of the disease in the sub-populations, or differential sensitivity to available therapies due to underlying tumour biology. Another factor of possible heterogeneity is the prognostic impact conferred by the biomarker itself. Furthermore, the same treatment acting on the same driver may trigger variable effects depending on differences in tumour biology, such as the presence of alternative drivers, escape pathways or resistance mechanisms. Central confirmation of biomarker status is highly recommended to ensure reliability of the results.

Umbrella trials

From a regulatory perspective, umbrella trials may be viewed essentially as a collection of parallel trials, each for a different compound or combination. More complex designs with possibly overlapping populations (e.g. multiple biomarkers expressed) and perhaps also a common control group on the other hand must be interpreted differently and might pose challenges from a statistical and regulatory perspective. It should be remembered that the allocation of different targeted therapies based on biomarker status in an umbrella study could result in treatment arms with different underlying prognosis, conferred by the biomarker, potentially hampering comparison with a common control arm.

Umbrella trials can serve exploratory purposes (e.g. to identify treatments for further development and to inform of activity), followed by standard confirmatory study designs. Umbrella trials may also serve as pivotal studies for market authorisation, when the general requirements for such are met, e.g. when randomised control arms are included, the power and Type I error control of studies are adequate. In

highly challenging situations and upon careful justification, uncontrolled umbrella trials might provide pivotal evidence and the methodological and regulatory principles and challenges will not differ from what is discussed for single-arm trials.

Basket trials

Basket designs can be used for different purposes with diametrically opposite objectives. They can be used for early phase trials aimed to identify patient populations likely to respond to the treatment for further development. In these cases, the objective is to detect differences in activity between baskets.

When basket trials are intended to serve as pivotal evidence for registration of a histology-independent indication, and when analyses across subpopulations ("pooling of baskets") are performed, there should be reassurance that there is no clear deviation from homogeneity of the treatment effect across baskets. A meaningful assessment of deviation from homogeneity is possible only if a sufficient number of patients from each subpopulation is included. This may not be generally feasible. Therefore, as there are limited possibilities to demonstrate homogeneity of the sub-populations in the baskets, a strong rationale to support a homogeneous treatment effect has to be provided upfront based on mechanistic rationale, pre-clinical data and pharmacodynamics, which need to be supported by the clinical data from the basket study. In particular, sponsors must justify and make it convincingly plausible by clinical and/or pre-clinical data that the interaction with tumour site or histology is negligible and this should also be supported by the final data.

A scenario for a basket development might include at least one relatively common sub-population, large enough to allow an acceptably powered controlled evidence for a relevant effect of the experimental treatment. Starting from this robust evidence, the conclusion could extend to smaller groups provided that expected or known heterogeneity between the sub-populations in the baskets does not prevent this extension exercise, a priori (see above) or *a posteriori* based on the trial results observed effects of the experimental treatment.

It should be noted that the relevance of an observed effect may differ importantly in relation to available therapeutic options for different disease entities.

9. Safety

In early stages of drug development as well as in the confirmatory setting used for regulatory benefit-risk assessment, the quality and informativeness of safety data is crucial.

9.1. Basic concepts

The concept of adverse drug reactions (ADRs) includes the implication of causality. In clinical trials, information on adverse events (AEs) with or without a causal relationship to the drug(s) should always be collected and graded by severity. Following causality assessment, some AEs will be determined to be ADRs. For an exact definition of what constitutes an ADR or AE, please refer to the ICH E2A guideline on clinical safety data management. In addition, the concept of treatment-emergent AEs (TEAEs) denotes AEs that were not present at baseline (pre-treatment) or have increased in severity grade during treatment (see ICH E9 guideline). The current standard grading system for AEs in oncology is the NCI CTCAE toxicity criteria. Tolerability may also be further addressed by using patient-reported outcomes (see Appendix 2).

The tolerability of a drug is often defined as the degree to which the adverse effects are acceptable to a patient. This suggests ADRs that affect the patient's quality of life or activities of daily living, often over a large proportion of the treatment time. In oncology these reactions typically include diarrhoea, mucositis, rash and neuropathy. This type of reactions may hamper the possibility of delivering the

drug at intended dose and schedule. Outcomes such as dose adjustments and discontinuation rate often provide important information on tolerability.

The importance of ADRs affecting tolerability versus infrequent severe or life-threatening ADRs differs depending on the disease setting. This needs to be considered in the planning of development programs.–Infrequent severe or even fatal ADRs may for example be considered an acceptable risk in the palliative setting if combined with good tolerability, while such a safety profile would make early neoadjuvant trials inappropriate.

9.2. Safety in the oncology context

In oncology the causality of adverse events in relation to the investigational drug is often difficult to assess due to overlapping symptoms of the underlying malignant disease and toxicity from backbone anticancer therapies, and the problem may be further emphasised by non-randomised study designs. This poses particular challenges to the understanding of an anticancer product's safety profile. Furthermore, it is not uncommon that certain adverse drug reactions are most prominent during the first to second treatment cycle(s), following which tolerance appears to develop. On the other hand, there is cumulative toxicity, of consequence mainly to those who have long-term treatment benefit. In these circumstances, cumulative ADR incidences alone do not sufficiently describe a product's safety profile.

The major groups of current pharmacological treatments include cytotoxins, targeted drugs, and immune modulators. The different dosing regimens and modes of action of these pharmacological and biological entities affect the toxicity and tolerability profiles in different ways, which must be considered in the planning of the collection, analysis and reporting of safety data. Conventional cytotoxic drugs are typically given at weekly or longer intervals and are characterised by major acute but transient toxicity, followed by recuperation before the next treatment cycle. In contrast, targeted drugs and immune modulators are typically administered continuously/daily, causing a different presentation of toxicities, including toxicities that are delayed or those that are more or less constant. For some products tolerability could be the major issue, while for others it can be potentially life-threatening adverse reactions. Both types of toxicity should be comprehensively investigated. The frequent coadministration of drugs from these major pharmacological groups further add to the complexity and demands on the safety collection and analysis.

In addition, there are advanced therapies, such as recombinant viral therapies and cell therapies whose particular safety profiles must be considered in the planning and reporting of studies.

9.3. Study design from a safety perspective

General recommendations

From a planning perspective it is important to consider how the study design impacts on the safety information obtained. General recommendations include the following.

In trials where the planned in-clinic treatment schedules differ between the randomised groups, the study design should aim to minimize differential surveillance, e.g. by phone-calls visits.

Assessment of safety from single-arm studies poses particular challenges as the lack of comparative data hampers the causality assessment. E.g. for haematology products it is not uncommon that many of the most frequently observed AEs are events that can be expected as symptoms of the underlying haematological malignancy, such as myelosuppression, infections, and bleeding. Therefore, whenever possible, comparative studies are recommended for marketing authorisation.

The need for post-authorisation generation of safety data should be considered prospectively, particularly if an early marketing authorisation is sought, e.g. conditional marketing authorisation.

For considerations regarding the definition of dose-limiting toxicities (DLTs) in the design of phase I studies depending on type of agent, please refer to section 6.2.1.

Extended safety data collection

A common problem with comparative studies is when the experimental drug shows substantially improved efficacy and patients therefore stay longer on the experimental arm than on the comparator arm. This introduces a bias by observation time if the collection of AEs is stopped at the time of study drug discontinuation or shortly thereafter. Furthermore, the "real-life" safety consequences of the comparator arm will be underestimated; both in the situation when there are no next-line therapies and the symptoms of disease increase after progression and discontinuation of study-drug, and when next-line therapies are administered with their consequent ADRs. Such post-therapy outcomes, particularly in the study arm with lower efficacy, can be of importance to the benefit-risk assessment by contextualising the risks of the experimental arm.

Extended safety data collection, including off-therapy and on-new therapy, may therefore be included in the study design, even if not chosen as the primary analysis cut-off for safety outcomes. In these designs, patients may not be discontinued from study at progression (unless enrolled in new study by a different sponsor with data exclusivity). This should be considered in particular when maintenance therapy is being investigated, in situations where analysis of PFS2 will be needed, or when the reversibility of an important ADR is of interest. PRO-measures may be of additional value in these situations. Depending on the situation, the specific rationale for extended safety monitoring may be used to define the appropriate scope (e.g. limited to specific ADRs) and appropriate duration of off-treatment safety data collection, in order to minimise burden on patients and impact on enrolment and compliance. The length of the extended safety data collection may also vary depending on the expected difference in time on treatment between study arms. The collection time should be sufficiently long to allow capture of both the increased symptomatology and decline in wellbeing associated with disease progression, as well as the ADRs of next-line therapy.

Safety database

The safety data base is comprised of all relevant studies and may include studies in other indications when extrapolation is justified. The size of the safety data base should be sufficient for benefit-risk assessment in the specific target population studied. The size required will depend on factors such as the severity of the sought indication and available treatment options, as well as on how large the benefit is. Even when a relatively small safety database is accepted at first approval, completion to full safety information is expected in a timely manner. Of note, when a treatment regimen is known to be associated with potentially fatal toxicity, such as high dose therapy in patients planned to undergo HSCT, this should normally be reflected in the choice of primary endpoint, i.e. whenever feasible overall survival, detailing treatment related mortality as predefined.

Demonstration of improved safety as study intent

Specific safety issues may sometimes be best addressed in dedicated studies. Such studies could be considered at any time during the development programme.

If the aims of a study include demonstration of improved safety, the protocol should specify how this should be accomplished, including with regard to sample size calculations. The often-non-constant rate (hazard) of toxicity events should be taken into account, both in the planning of the study and in the analysis of study data (see further below).

It is not acceptable to focus on one toxic effect only. In addition to a specific item, such as neuropathy, where a clinically relevant improvement is expected, the outcome measure(s) should provide unbiased information on overall toxicity and tolerability.

9.4. Safety data collection, analysis and reporting

All toxicity should be described, including cumulative toxicity. Exclusion of assumed disease-related events from collected data, even if based on reasonable assumptions, may hamper the ability of detecting a relationship (also) with the drug, and is therefore not allowed. If cure is the objective, long term follow-up for toxicity is highly relevant. Late toxicity typically occurs several years after treatment and includes second primary malignancies and certain organ toxicities (e.g. CNS, cardiovascular). The number of patients suffering from late toxicities may increase over time and is therefore an objective for post-licensure pharmacovigilance activities.

All marketing authorisation applications should include cumulative adverse event rates from the pivotal study(ies) at the specified time points 3 months, 6 months and 1 year, in order to facilitate regulatory safety assessment. In cases where the time on therapy is significantly shorter or longer, additional or alternative time-points (e.g. 1 month, 5 years) should be considered.

It is furthermore recommended that AEs leading to dose reduction, interruption and discontinuation are reported by relatedness according to the investigator. Laboratory abnormalities such as cytopenias or liver enzymes that lead to dose changes or -interruption should also preferably be reported with the summary term laboratory AEs.

Temporal perspective

In addition to standard reporting of adverse events based on cumulative frequencies by toxicity grade, complementary measurements are required for a thorough understanding of the safety profile of a given anticancer drug. It is important to understand how the incidence, prevalence and severity of certain AEs change with time on treatment.

For key events, i.e. events that are common and affect tolerability, safety by treatment cycle is often of value. For example, fatigue or diarrhoea grade 3 for limited periods of time may not affect tolerability to a great degree, while long-term fatigue or diarrhoea grade 2 may be a major issue to the benefit-risk balance and may thus motivate specific analysis. Measurements such as incidence and prevalence per period of time or per treatment cycle, time to event, and duration of event (including by grade) should normally be considered. Patient-reported outcomes may also be useful in the evaluation (see Appendix 2).

Time-adjusted analyses for AEs, e.g. incidence by different cut-off dates or event rates per 100 patient-years, may also be indicated if properly justified by the pattern of events. While the rate of events may seldom be constant, thus precluding formal statistical comparison of the raw event rates (which would require the assumption of exponential distribution), such descriptive summaries often facilitate the assessment when the observation time differs importantly across study arms. In addition, Kaplan-Meier analysis of selected AEs, which considers censoring of events, may be useful. Not all AEs may need to be reported in such detail, however. Selection criteria can for example include events leading to dose withdrawal, reduction or interruption, serious adverse events, and events that are likely to affect tolerability or the benefit-risk balance.

Dose reductions and other consequences

To what extent dose reductions alleviate the event(s) that lead to dose reduction in the first place may be of importance to the benefit-risk assessment. It is expected that the use and effects of preventive measures, such as anti-emetics or growth factors are reported.

Understanding relationship between the AE profile and drug exposure might be of importance. In addition, longitudinal PK/PD-data, where dose adjustments are taken into account, may provide further insights.

Additional characterisation of the consequences of ADRs may sometimes be warranted, e.g. severity and type of infections associated with neutropenia, hospitalisation rates and duration, resource utilisation (e.g. transfusions) and outcomes including recovery and fatality rates.

Monitoring of frequency and type (viral, bacterial, fungal) of possible, probable or proven infections should be undertaken in patients undergoing more intensive cytotoxic/immunosuppressive therapy. For compounds known or suspected to cause long term immunodeficiency, monitoring for opportunistic infections for up to one year after the end of therapy should be considered.

For immunomodulatory agents such as checkpoint inhibitors, awareness and monitoring of potential development of immune-related adverse events such as diarrhoea/colitis, rash, mucositis, liver toxicity, hypophysitis, pneumonitis and other endocrinopathies are important.

Causality assessment

Causality assessment is a critical step in establishing a safety profile. In oncology, this may present particular challenges, as discussed above (Section 8.2) The principles for causality assessment outlined in the SmPC guideline should be adhered to. In addition, the following should be considered.

- Care should be taken in order not to dilute the product information with unrelated AEs.
- The conclusion of which AEs constitute ADRs should not rely solely on the investigator assessments of causality.

While the investigator causality assessments of individual patients may not be changed and must be presented as reported, the applicant of a marketing authorisation is responsible for the establishment and communication of the product's safety profile, which should be based on a thorough evaluation of the (preclinical and clinical) safety data.

This is motivated partly by the fact that when the pivotal studies used for the first marketing authorisation approval are performed, the knowledge of the product's true safety profile is limited. The investigator assessments of adverse events' relatedness to study drug may therefore be more prone to error in these first studies compared with studies of approved drugs, in particular for events that are overlapping with the symptoms of the disease or otherwise expected in the patient population. For these, relatedness to study drug may tend to be underestimated. In other situations, investigators may overestimate the relatedness. Thus, while investigator assessments of causality may often provide useful clinical insights, the all-causality AE frequencies may be expected to be the measure least biased by preformed understanding.

In situations where a sufficiently large and undisputed difference in AE frequency between randomised study arms is not present to base the conclusion of an ADR on, the Sponsor's causality assessment must include a medical-pharmacological assessment. In the absence of a clear known pharmacological mechanism, factors making a causal relationship plausible, such as positive de-challenge and re-challenge, should be taken into account. In cases with AEs deemed as (possibly) related by investigators, but containing too limited information to allow secondary assessment of causality by the Sponsor and regulatory authorities, all efforts should be made to procure more data. Standardised MedDRA Queries (SMQs) including broad terms may provide additional insight. If the lack of data persists, an ADR should not be concluded until sufficiently informative cases have occurred.

Oncology drugs are frequently administered in combinations. Irrespective of design, e.g. BA vs. A or BA vs. CA, it may not be possible to define causality in relation to the individual drugs. These attempts should not overshadow the main objective, i.e. to define causality of AEs in relation to the regimens under study.

9.5. Laboratory abnormalities

While laboratory abnormalities reported as AEs might be interpreted as those that were perceived by investigators to be clinically relevant, the unbiased registration of laboratory values from clinical trials is considered a more reliable measure. Both types of data can provide valuable information, but the risk of bias in investigator reports of laboratory AEs should be considered. As with other TEAEs, longitudinal analysis, including impact of dose adjustments, and time-dependent analyses may be of value.

Baseline factors that may affect the causality assessment with regard to treatment-emergent laboratory abnormalities should also be considered and additional analyses may be required to assess causality. For example, if a large proportion of the patients in the study population have baseline liver metastases it is unlikely that the total frequency of liver enzyme elevations is caused by the drug. In these situations, additional separate analyses may be employed for patients with and without confounding factors, such as liver metastases in this case.

9.6. Safety issues related to radiation therapy

As radiation therapy is a standard treatment option in many malignant tumours, it is foreseeable that patients will be receiving radiation therapy. Information on concomitant or sequential use of the medicinal agent with radiotherapy should therefore be collected throughout the entire study programme, including data on dose, fraction, target/field and time. The safety data collection and reporting should address radiotherapy specific items such as radio-sensitisation and "radiation recall". The detailed information on the administered radiotherapy may be crucial to the possibility to understand in retrospect unforeseen radio sensitisation reactions when they occur, and to give recommendations for precautions. Subjects requiring radiation therapy due to progressive disease while enrolled in a trial of a novel agent or combination of agents will normally be withdrawn from study therapy, as progression is usually a stopping rule, unless the study design includes other predefined measures to handle such events.

9.7. Using patient reported outcomes in the safety assessment

Patient reported outcomes (PROs) may be complementary tools for assessing the tolerability of anticancer products' safety profiles, including in the evaluation of the effect of dose-reductions on ADRs. (See PRO appendix to this guideline.)

9.8. Safety reporting in special populations and pharmacogenomics

It is recommended that samples are collected prospectively to enable pharmacogenomic evaluation in relation to safety issues, as appropriate.

Safety in special populations, as detailed above (Sections 4 and 7.7), should be summarised from the full studies programme.

Paediatric population

For studies in the paediatric population, adverse events should include the reporting of any observed effects on organ maturation, growth and development, including fertility. Some of these long-term aspects will require further follow-up in the post authorisation setting, while non-clinical studies may provide an important source of information for the benefit-risk assessment at the time of market authorisation.

Other important issues for evaluation in paediatric studies may include whether the toxicity profile and/or its impact differ compared with adults or between different paediatric age groups. The difference in robustness when comparing data sets of markedly different sizes (e.g. adult vs. paediatric population) should be taken into account. While adequate empirical comparative data form the basis of the safety evaluation, modelling and simulations may provide complementary information where data in (parts of) the paediatric population are difficult to obtain.

Elderly patients and other risk factors

Registration studies should aim to include elderly or frail patients if these are expected to be part of the target population. The safety profile in these subgroups should be reported.

Similarly, if foreseen to be treated with the drug when authorised, patients with risk factors such as poor performance status or brain metastasis should be included whenever possible in order to generate

safety data in these subgroups, of relevance to the future prescribing information. They may, however, be excluded from the primary analysis population, as regards to both efficacy and safety.

9.9. Presentation of adverse drug reactions in the product information

In oncology, symptoms of the disease may be prominent and indistinguishable from the corresponding drug reaction (e.g. fatigue, weight loss, gastrointestinal symptoms, and myelosuppression – depending on the disease). Similarly, it may be impossible to determine the contribution of toxicity from different agents when combination therapy is given. This makes communication of drug toxicity to the prescriber and patient challenging. To address such situations, and in order to achieve consistency and comparability across the SmPCs of different products, the following practical recommendations should be considered together with the principles described in the SmPC guideline on section 4.8. (see also Appendix: The SmPC for Anticancer medicinal products)

As there is often no way to identify the "true" incidence of an ADR, the least biased measure should be consistently used. For events fulfilling the causality requirement of ADR, the frequency categories in the tabulated list of adverse reactions should therefore be based on the frequencies of all-causality AEs (i.e. irrespective of investigators' assessments of relatedness). It should be clearly communicated in the SmPC, however, that the ADR frequencies presented may not be fully attributable to the drug alone but may contain contributions from the underlying disease or from other drugs used in a combination. In addition, the median observation time upon which the ADR frequencies are based should be given in the SmPC Section 4.8 for contextualisation and to facilitate across-product comparisons. Information on frequencies by toxicity grade is often of value to the prescriber and should normally be included for toxic anticancer agents, e.g. reactions of all grades compared with grade ≥3.

Comparative data, i.e. information from the control arm in randomised studies, may be presented for selected reactions of interest for contextualisation. Selection criteria may include e.g. those leading to discontinuation, dose reduction or interruption, serious adverse reactions, and reactions that are likely to affect tolerability or the benefit-risk balance. This information may be placed after the main ADR table in SmPC Section 4.8 (subsection c). If justified, data from several trials may be presented separately (e.g. to allow comparison of incidences in studies with different designs). However, when resulting in a more accurate and reliable estimation, pooled analysis across suitable studies will be preferred also for readability purposes.

Presentation of information on additional informative measures discussed above may also be warranted (e.g. duration of selected ADRs, time-adjusted ADR frequencies etc.)

For laboratory abnormalities, data from the unbiased collection of laboratory data should normally be presented in the SmPC and may also be complemented by comparative data when justified.

Definitions and abbreviations

ADCC: Antibody-dependent cell-mediated cytotoxicity

ADR: Adverse drug reaction

AE: Adverse event

ANC: Absolute neutrophil count

BSA: Body surface area

BSC: Best supportive care – include antibiotics, nutritional support, correction of metabolic disorders, optimal symptom control and pain management (including radiotherapy), etc. but does not include tumour specific therapy

CBR: Clinical benefit rate; also, Clinical benefit response. CR or PR or prolonged SD. "Prolonged SD" is defined condition specific, for breast cancer normally ≥24 weeks.

Chemoprotectant: A compound which counteracts the activity of anti-tumour compounds on normal tissue without (or clearly less) affecting the anti-tumour activity.

Chemosensitizer (or drug resistance modifier): A compound without own anti-tumour activity which increases the activity through pharmacodynamic interaction with anti-tumour compound(s).

Cytostatic: Anticancer compound shown to inhibit cell division without direct effects on tumour cell viability in non-clinical studies.

Cytotoxic: Anticancer compounds inducing irreversible lethal lesions through interference with DNA replication, mitosis, etc. following short term exposure in non-clinical studies.

CR: Complete response

CRF: Case report form

CTLA-4: Cytotoxic T-lymphocyte-associated protein 4

Data maturity: A clinical study is considered mature if the distribution of events over time (early – late) makes it feasible to estimate the treatment effect in the full study population. This refers to the assumption that there is a biological difference between e.g. tumours progressing early and late and that the treatment effect might differ. The number of late events should therefore be large enough for study data to be stable. In practice, if a treatment difference has been established and a clear majority of events expected over long term have occurred, the study may in most cases be regarded as "mature".

DFS: Disease-free survival (time from randomisation to recurrence or death from any cause) **DLT:**

Dose limiting toxicities

EFS: Event-free survival in this guideline refers to lack of achievement of CR, relapse and death without relapse are counted as events in an EFS analysis. Those patients who did not reach CR during the pre-specified induction phase will be considered as having an event at time 0.

FcRn: The neonatal Fc receptor

HRQoL: Health related quality of life

IgG: Immunoglobulin G

MedDRA: Medical Dictionary for Regulatory Activities

MoAb: Monoclonal antibody

MTA: Molecularly targeted agents

MTD: Maximum tolerated dose; the highest dose of drug that can be given without causing unacceptable adverse reactions in most recipients. Determined in phase I-studies, the MTD has

traditionally often been defined by dose-limiting toxicity occurring in at least 2 of 6 patients so that further dose-escalation is not undertaken. Other definitions and algorithms are also used.

NCI: National Cancer Institute

Non-cytotoxic: Anticancer compounds not belonging to the class of cytotoxic compounds.

ORR: Objective response rate (the proportion of patients in whom a CR or PR was observed)

OS: Overall survival (time from randomisation to death from any cause)

PD: Pharmacodynamics

PD-1: Programmed death-1 receptor

PD-L1: Programmed death-ligand 1

PK: Pharmacokinetics **PR**:

Partial response

Primary (innate) resistance: Progression without prior objective response or growth inhibition.

PRO: Patient reported outcome

PFS: Progression-free survival (time from randomisation to objective tumour progression or death from any cause)

PFS2: PFS on next-line therapy. Time from randomisation to objective tumour progression on nextline treatment or death from any cause. In some cases, time on next line therapy may be used as proxy for PFS2.

QoL: Quality of life

Randomised phase II trial: Randomised exploratory study designed to provide data of importance for the design of Phase III confirmatory studies, e.g. with respect an estimate of the possible magnitude of the effect using a clinically relevant measure of activity and/or biomarkers.

Refractory: Progression on therapy or within a short period of time after last cycle of therapy.

Resistance: Progression within a defined timeframe after end of therapy.

RP2D: Recommended phase 2 dose

SD: Stable disease

Secondary resistance: Progression after documented objective response or period of growth inhibition.

SMQ: Standard MedDRA queries

TEAE: Treatment-emergent adverse event. An event that emerges during treatment having been absent pre-treatment or worsens relative to the pre-treatment state. (See ICH E9)

TTF: Time to treatment failure (time from randomisation to discontinuation of therapy for any reason including death, progression, toxicity or add-on of new anti-cancer therapy)

TTP: Time to tumour progression (time from randomisation to observed tumour progression, censoring for death not related to the underlying malignancy)

Window of opportunity: Under certain well-defined conditions it is acceptable to conduct a clinical study with an experimental compound in settings (line of therapy, stage, etc.) where available data for this compound normally would be regarded as too limited. The conditions for conducting such a study must be set rigorously so that the interest of the patient is guaranteed. Circumstances to take into account include benefit-risk of available therapies, available safety/activity data for the experimental compound, tumour-related symptoms (in most cases absent), expected evolution of the disease if left

untreated or treated with available therapies, ease of frequent monitoring of tumour evolution (including use of biomarkers), planned intervention post chemotherapy, etc.