Food and Drug Administration
Center for Drug Evaluation and Research

Oncologic Drugs Advisory Committee Meeting
March 13, 2006
Briefing Material – Morning Session

Table of Contents

iii) General Guide for Starting Dose Selection for a Cytotoxic Agent in Cancer Patients (Flowchart)(http://www.fda.gov/cder/cancer/docs/doseflow.pdf)

c. Horstmann E, et al (2005) Risks and Benefits of Phase 1 Oncology Trials, 1991 through 2002. N Engl J Med 352 (9) 895-904. (http://content.nejm.org/cgi/reprint/352/9/895.pdf)

4. Response Evaluation Criteria in Solid Tumors (RECIST): Standard parameters to be used when documenting response for solid tumors. (http://ctep.cancer.gov/guidelines/recist.html)

a. Therasse P, et al. New Guidelines to Evaluate the Response to Treatment in Solid Tumors. J Natl Cancer Inst 2000; 92: 205-216. (http://jncicancerspectrum.oxfordjournals.org/cgi/reprint/jnci;92/3/205.pdf)

b. RECIST Quick Reference (from NCI/CTEP website http://ctep.cancer.gov/guidelines/recist.html)

DEPARTMENT OF HEALTH & HUMAN SERVICES

Public Health Service

Food and Drug Administration
Rockville, MD 20857

Memorandum

Date: February 14, 2006

To: Oncologic Drugs Advisory Committee (ODAC) Members and Guests

From: Richard Pazdur, M.D.
Director, Office of Oncology Drug Products, CDER, FDA

Subject: FDA Background Package for March 13, 2006 ODAC Meeting
on Nonclinical Study Requirements for Oncology Drugs and Biologics

---

This memo outlines the purpose of the morning session of the March 13, 2006 ODAC meeting. This session will provide the Committee with an overview of the current FDA requirements for nonclinical safety evaluation of new small molecule and biotechnology-derived drugs prior to initial use in human subjects, in the context of development of products for the treatment of cancer.

Applicants submitting Investigational New Drug Applications (INDs) to the FDA for early clinical investigations of new biological or small molecule drugs are required to include data from nonclinical animal and/or *in vitro* pharmacology and toxicology studies. This requirement is derived from the Federal Food Drug and Cosmetic Act of 1938 (FD&C Act), and codified in the Code of Federal Regulations [21 CFR 312.23(8)]. The data resulting from these studies provide the basis for which the sponsor and ultimately the FDA, conclude that the product is reasonably safe for clinical use.

Although the nonclinical pharmacology and toxicology studies are critical to demonstrate the safety and support the rationale for the clinical investigation, the kind, duration, and scope of animal and other safety testing varies with the duration and nature of the proposed clinical use. The FDA recognizes that unique issues arise in designing and interpreting nonclinical studies for small molecule drugs and biological therapeutics, and has provided several guidance documents to assist investigators in developing their nonclinical programs. Guidance documents are also available through the International Conference on Harmonisation (ICH) that provide a framework for the design of nonclinical safety studies, with the objective of adequately meeting the requirements set forth by the FDA and other, global regulatory authorities. However, these documents deliberately do not provide a universal nonclinical study paradigm by which all investigational drugs and biologics may be tested, since flexibility is needed to address specific concerns related to the biology of the product itself, the clinical population planned for study, and the indication under investigation.

With the recent reorganization of the structure of the Office of New Drugs at FDA, the 2003 transfer of the biologic therapeutic products from CBER to CDER, and the creation

of the Office of Oncology Drug Products at FDA in 2005, the differences in the nonclinical testing requirements for the different product classes have been recognized by both industry and the FDA. The Office of Oncology Drug Products is currently working to develop guidance for nonclinical testing standards for new molecular entities, including both small molecular weight drugs and biotechnology-derived therapeutic agents, with the goals of both identifying differences between the two product classes and providing appropriate guidance where necessary, and harmonizing nonclinical testing standards where appropriate.

FDA requests that ODAC discuss the written questions provided in the briefing package, and offer advice to the Agency on how adequate nonclinical safety data to support early clinical investigations in oncology may be obtained, without adding undue burden to sponsors for their clinical development programs. The goal of this discussion is for recommendations made by the Committee today to be incorporated into the FDA guidance, to facilitate new approaches to the treatment of cancer.

Guidance for Industry

S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals

July 1997
ICH

Guidance for Industry

S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals

Additional copies are available from:
the Drug Information Branch (HFD-210),
Center for Drug Evaluation and Research (CDER),
5600 Fishers Lane, Rockville, MD 20857 (Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm
or
Office of Communication,
Training, and Manufacturers Assistance (HFM-40)
Center for Biologics Evaluation and Research (CBER)
1401 Rockville Pike, Rockville, MD 20852-1448,
http://www.fda.gov/cber/guidelines.htm
(Fax) 888-CBERFAX or 301-827-3844
(Voice Information) 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
July 1997
ICH

Table of Contents

GUIDANCE FOR INDUSTRY[1]

S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals

I. INTRODUCTION (1)

A. Background (1.1)

Biotechnology-derived pharmaceuticals (biopharmaceuticals) were initially developed in the early 1980's. The first marketing authorizations were granted later in the decade. Several guidelines and points-to-consider documents have been issued by various regulatory agencies regarding safety assessment of these products. Review of such documents, which are available from regulatory authorities, may provide useful background in developing new biopharmaceuticals.

Considerable experience has now been gathered with submission of applications for biopharmaceuticals. Critical review of this experience has been the basis for development of this guidance, which is intended to provide general principles for designing scientifically acceptable preclinical safety evaluation programs.

B. Objectives (1.2)

Regulatory standards for biotechnology-derived pharmaceuticals have generally been comparable among the European Union, Japan, and the United States. All three regions have adopted a flexible, case-by-case, science-based approach to preclinical safety evaluation needed to support clinical development and marketing authorization. In this rapidly evolving scientific area, there is a need for common understanding and continuing dialogue among the regions.

The primary goals of preclinical safety evaluation are: (1) To identify an initial safe dose and subsequent dose escalation schemes in humans; (2) to identify potential target organs for toxicity and for the study of whether such toxicity is reversible; and (3) to identify safety parameters for clinical monitoring. Adherence to the principles presented in this document should improve the

---

[1] This guidance was developed within the Expert Working Group (Safety) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, July 1997. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and the United States. This guidance was published in the *Federal Register* on November 18, 1997 (62 FR 61515), and is applicable to drug and biological products. This guidance represents the Agency's current thinking on preclinical safety evaulation of biotechnology-derived pharmaceuticals. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

quality and consistency of the preclinical safety data supporting the development of biopharmaceuticals.

C. Scope (1.3)

This guidance is intended primarily to recommend a basic framework for the preclinical safety evaluation of biotechnology-derived pharmaceuticals. It applies to products derived from characterized cells through the use of a variety of expression systems including bacteria, yeast, insect, plant, and mammalian cells. The intended indications may include in vivo diagnostic, therapeutic, or prophylactic uses. The active substances include proteins and peptides, their derivatives, and products of which they are components; they could be derived from cell cultures or produced using recombinant deoxyribonucleic acid (DNA) technology, including production by transgenic plants and animals. Examples include but are not limited to: Cytokines, plasminogen activators, recombinant plasma factors, growth factors, fusion proteins, enzymes, receptors, hormones, and monoclonal antibodies.

The principles outlined in this guidance may also be applicable to recombinant DNA protein vaccines, chemically synthesized peptides, plasma derived products, endogenous proteins extracted from human tissue, and oligonucleotide drugs.

This document does not cover antibiotics, allergenic extracts, heparin, vitamins, cellular blood components, conventional bacterial or viral vaccines, DNA vaccines, or cellular and gene therapies.

II. SPECIFICATION OF THE TEST MATERIAL (2)

Safety concerns may arise from the presence of impurities or contaminants. It is preferable to rely on purification processes to remove impurities and contaminants rather than to establish a preclinical testing program for their qualification. In all cases, the product should be sufficiently characterized to allow an appropriate design of preclinical safety studies.

There are potential risks associated with host cell contaminants derived from bacteria, yeast, insect, plants, and mammalian cells. The presence of cellular host contaminants can result in allergic reactions and other immunopathological effects. The adverse effects associated with nucleic acid contaminants are theoretical but include potential integration into the host genome. For products derived from insect, plant, and mammalian cells, or transgenic plants and animals, there may be an additional risk of viral infections.

In general, the product that is used in the definitive pharmacology and toxicology studies should be comparable to the product proposed for the initial clinical studies. However, it is appreciated that during the course of development programs, changes normally occur in the manufacturing

process in order to improve product quality and yields. The potential impact of such changes for extrapolation of the animal findings to humans should be considered.

The comparability of the test material during a development program should be demonstrated when a new or modified manufacturing process is developed or other significant changes in the product or formulation are made in an ongoing development program. Comparability can be evaluated on the basis of biochemical and biological characterization (i.e., identity, purity, stability, and potency). In some cases, additional studies may be needed (i.e., pharmacokinetics, pharmacodynamics and/or safety). The scientific rationale for the approach taken should be provided.

III. PRECLINICAL SAFETY TESTING (3)

A. General Principles (3.1)

The objectives of the preclinical safety studies are to define pharmacological and toxicological effects not only prior to initiation of human studies but throughout clinical development. Both in vitro and in vivo studies can contribute to this characterization. Biopharmaceuticals that are structurally and pharmacologically comparable to a product for which there is wide experience in clinical practice may need less extensive toxicity testing.

Preclinical safety testing should consider: (1) Selection of the relevant animal species; (2) age; (3) physiological state; (4) the manner of delivery, including dose, route of administration, and treatment regimen; and (5) stability of the test material under the conditions of use.

Toxicity studies are expected to be performed in compliance with Good Laboratory Practice (GLP); however, it is recognized that some studies employing specialized test systems, which are often needed for biopharmaceuticals, may not be able to comply fully with GLP. Areas of noncompliance should be identified and their significance evaluated relative to the overall safety assessment. In some cases, lack of full GLP compliance does not necessarily mean that the data from these studies cannot be used to support clinical trials and marketing authorizations.

Conventional approaches to toxicity testing of pharmaceuticals may not be appropriate for biopharmaceuticals due to the unique and diverse structural and biological properties of the latter that may include species specificity, immunogenicity, and unpredicted pleiotropic activities.

B. Biological Activity/Pharmacodynamics (3.2)

Biological activity may be evaluated using in vitro assays to determine which effects of the product may be related to clinical activity. The use of cell lines and/or primary cell cultures can be useful to examine the direct effects on cellular phenotype and proliferation. Due to the species specificity of many biotechnology-derived pharmaceuticals, it is important to select relevant

animal species for toxicity testing. In vitro cell lines derived from mammalian cells can be used to predict specific aspects of in vivo activity and to assess quantitatively the relative sensitivity of various species (including human) to the biopharmaceutical. Such studies may be designed to determine, for example, receptor occupancy, receptor affinity, and/or pharmacological effects, and to assist in the selection of an appropriate animal species for further in vivo pharmacology and toxicology studies. The combined results from in vitro and in vivo studies assist in the extrapolation of the findings to humans. In vivo studies to assess pharmacological activity, including defining mechanism(s) of action, are often used to support the rationale of the proposed use of the product in clinical studies.

For monoclonal antibodies, the immunological properties of the antibody should be described in detail, including its antigenic specificity, complement binding, and any unintentional reactivity and/or cytotoxicity towards human tissues distinct from the intended target. Such cross-reactivity studies should be carried out by appropriate immunohistochemical procedures using a range of human tissues.

C. Animal Species/Model Selection (3.3)

The biological activity together with species and/or tissue specificity of many biotechnology-derived pharmaceuticals often preclude standard toxicity testing designs in commonly used species (e.g., rats and dogs). Safety evaluation programs should include the use of relevant species. A relevant species is one in which the test material is pharmacologically active due to the expression of the receptor or an epitope (in the case of monoclonal antibodies). A variety of techniques (e.g., immunochemical or functional tests) can be used to identify a relevant species. Knowledge of receptor/epitope distribution can provide greater understanding of potential in vivo toxicity.

Relevant animal species for testing of monoclonal antibodies are those that express the desired epitope and demonstrate a similar tissue cross-reactivity profile as for human tissues. This would optimize the ability to evaluate toxicity arising from the binding to the epitope and any unintentional tissue cross-reactivity. An animal species that does not express the desired epitope may still be of some relevance for assessing toxicity if comparable unintentional tissue cross-reactivity to humans is demonstrated.

Safety evaluation programs should normally include two relevant species. However, in certain justified cases one relevant species may suffice (e.g., when only one relevant species can be identified or where the biological activity of the biopharmaceutical is well understood). In addition, even where two species may be necessary to characterize toxicity in short term studies, it may be possible to justify the use of only one species for subsequent long-term toxicity studies (e.g., if the toxicity profile in the two species is comparable in the short term).

Toxicity studies in nonrelevant species may be misleading and are discouraged. When no relevant species exists, the use of relevant transgenic animals expressing the human receptor or the use of

homologous proteins should be considered. The information gained from use of a transgenic animal model expressing the human receptor is optimized when the interaction of the product and the humanized receptor has similar physiological consequences to those expected in humans. While useful information may also be gained from the use of homologous proteins, it should be noted that the production process, range of impurities/contaminants, pharmacokinetics, and exact pharmacological mechanism(s) may differ between the homologous form and the product intended for clinical use. Where it is not possible to use transgenic animal models or homologous proteins, it may still be prudent to assess some aspects of potential toxicity in a limited toxicity evaluation in a single species, e.g., a repeated dose toxicity study of $\leq$ 14 days duration that includes an evaluation of important functional endpoints (e.g., cardiovascular and respiratory).

In recent years, there has been much progress in the development of animal models that are thought to be similar to the human disease. These animal models include induced and spontaneous models of disease, gene knockout(s), and transgenic animals. These models may provide further insight, not only in determining the pharmacological action of the product, pharmacokinetics, and dosimetry, but may also be useful in the determination of safety (e.g., evaluation of undesirable promotion of disease progression). In certain cases, studies performed in animal models of disease may be used as an acceptable alternative to toxicity studies in normal animals (Note 1). The scientific justification for the use of these animal models of disease to support safety should be provided.

D. Number/Gender of Animals (3.4)

The number of animals used per dose has a direct bearing on the ability to detect toxicity. A small sample size may lead to failure to observe toxic events due to observed frequency alone regardless of severity. The limitations that are imposed by sample size, as often is the case for nonhuman primate studies, may be in part compensated by increasing the frequency and duration of monitoring. Both genders should generally be used or justification given for specific omissions.

E. Administration/Dose Selection (3.5)

The route and frequency of administration should be as close as possible to that proposed for clinical use. Consideration should be given to pharmacokinetics and bioavailability of the product in the species being used and to the volume which can be safely and humanely administered to the test animals. For example, the frequency of administration in laboratory animals may be increased compared to the proposed schedule for the human clinical studies in order to compensate for faster clearance rates or low solubility of the active ingredient. In these cases, the level of exposure of the test animal relative to the clinical exposure should be defined. Consideration should also be given to the effects of volume, concentration, formulation, and site of administration. The use of routes of administration other than those used clinically may be acceptable if the route must be modified due to limited bioavailability, limitations due to the route of administration, or to size/physiology of the animal species.

Dosage levels should be selected to provide information on a dose-response relationship, including a toxic dose and a no observed adverse effect level (NOAEL). For some classes of products with little to no toxicity, it may not be possible to define a specific maximum dose. In these cases, a scientific justification of the rationale for the dose selection and projected multiples of human exposure should be provided. To justify high dose selection, consideration should be given to the expected pharmacological/physiological effects, availability of suitable test material, and the intended clinical use. Where a product has a lower affinity to or potency in the cells of the selected species than in human cells, testing of higher doses may be important. The multiples of the human dose that are needed to determine adequate safety margins may vary with each class of biotechnology-derived pharmaceutical and its clinical indication(s).

F. Immunogenicity (3.6)

Many biotechnology-derived pharmaceuticals intended for humans are immunogenic in animals. Therefore, measurement of antibodies associated with administration of these types of products should be performed when conducting repeated dose toxicity studies in order to aid in the interpretation of these studies. Antibody responses should be characterized (e.g., titer, number of responding animals, neutralizing or non-neutralizing) and their appearance should be correlated with any pharmacological and/or toxicological changes. Specifically, the effects of antibody formation on pharmacokinetic/pharmacodynamic parameters, incidence and/or severity of adverse effects, complement activation, or the emergence of new toxic effects should be considered when interpreting the data. Attention should also be paid to the evaluation of possible pathological changes related to immune complex formation and deposition.

The detection of antibodies should not be the sole criterion for the early termination of a preclinical safety study or modification in the duration of the study design unless the immune response neutralizes the pharmacological and/or toxicological effects of the biopharmaceutical in a large proportion of the animals. In most cases, the immune response to biopharmaceuticals is variable, like that observed in humans. If the interpretation of the data from the safety study is not compromised by these issues, then no special significance should be ascribed to the antibody response.

The induction of antibody formation in animals is not predictive of a potential for antibody formation in humans. Humans may develop serum antibodies against humanized proteins, and frequently the therapeutic response persists in their presence. The occurrence of severe anaphylactic responses to recombinant proteins is rare in humans. In this regard, the results of guinea pig anaphylaxis tests, which are generally positive for protein products, are not predictive for reactions in humans; therefore, such studies are considered of little value for the routine evaluation of these types of products.

IV. SPECIFIC CONSIDERATIONS (4)

A. Safety Pharmacology (4.1)

It is important to investigate the potential for undesirable pharmacological activity in appropriate animal models and, where necessary, to incorporate particular monitoring for these activities in the toxicity studies and/or clinical studies. Safety pharmacology studies measure functional indices of potential toxicity. These functional indices may be investigated in separate studies or incorporated in the design of toxicity studies. The aim of the safety pharmacology studies should be to reveal any functional effects on the major physiological systems (e.g., cardiovascular, respiratory, renal, and central nervous systems). Investigations may also include the use of isolated organs or other test systems not involving intact animals. All of these studies may allow for a mechanistically-based explanation of specific organ toxicities, which should be considered carefully with respect to human use and indication(s).

B. Exposure Assessment (4.2)

1. Pharmacokinetics and Toxicokinetics (4.2.1)

It is difficult to establish uniform guidances for pharmacokinetic studies for biotechnology-derived pharmaceuticals. Single and multiple dose pharmacokinetics, toxicokinetics, and tissue distribution studies in relevant species are useful; however, routine studies that attempt to assess mass balance are not useful. Differences in pharmacokinetics among animal species may have a significant impact on the predictiveness of animal studies or on the assessment of dose-response relationships in toxicity studies. Alterations in the pharmacokinetic profile due to immune-mediated clearance mechanisms may affect the kinetic profiles and the interpretation of the toxicity data. For some products, there may also be inherent, significant delays in the expression of pharmacodynamic effects relative to the pharmacokinetic profile (e.g., cytokines) or there may be prolonged expression of pharmacodynamic effects relative to plasma levels.

Pharmacokinetic studies should, whenever possible, utilize preparations that are representative of those intended for toxicity testing and clinical use and employ a route of administration that is relevant to the anticipated clinical studies. Patterns of absorption may be influenced by formulation, concentration, site, and/or volume. Whenever possible, systemic exposure should be monitored during the toxicity studies.

When using radiolabeled proteins, it is important to show that the radiolabeled test material maintains activity and biological properties equivalent to that of the unlabeled material. Tissue concentrations of radioactivity and/or autoradiography data using radiolabeled proteins may be difficult to interpret due to rapid in vivo metabolism or unstable radiolabeled linkage. Care should be taken in the interpretation of studies using

radioactive tracers incorporated into specific amino acids because of recycling of amino acids into nondrug related proteins/peptides.

Some information on absorption, disposition, and clearance in relevant animal models should be available prior to clinical studies in order to predict margins of safety based upon exposure and dose.

2. Assays (4.2.2)

The use of one or more assay methods should be addressed on a case-by-case basis and the scientific rationale should be provided. One validated method is usually considered sufficient. For example, quantitation of TCA-precipitable radioactivity following administration of a radiolabeled protein may provide adequate information, but a specific assay for the analyte is preferred. Ideally, the assay methods should be the same for animals and humans. The possible influence of plasma binding proteins and/or antibodies in plasma/serum on the assay performance should be determined.

3 Metabolism (4.2.3)

The expected consequence of metabolism of biotechnology-derived pharmaceuticals is the degradation to small peptides and individual amino acids. Therefore, the metabolic pathways are generally understood. Classical biotransformation studies as performed for pharmaceuticals are not needed.

Understanding the behavior of the biopharmaceutical in the biologic matrix (e.g., plasma, serum, cerebral spinal fluid) and the possible influence of binding proteins is important for understanding the pharmacodynamic effect.

C. Single Dose Toxicity Studies (4.3)

Single dose studies may generate useful data to describe the relationship of dose to systemic and/or local toxicity. These data can be used to select doses for repeated dose toxicity studies. Information on dose-response relationships may be gathered through the conduct of a single dose toxicity study or as a component of pharmacology or animal model efficacy studies. The incorporation of safety pharmacology parameters in the design of these studies should be considered.

D. Repeated Dose Toxicity Studies (4.4)

For consideration of the selection of animal species for repeated dose studies, see section III.C (3.3). The route and dosing regimen (e.g., daily versus intermittent dosing) should reflect the intended clinical use or exposure. When feasible, these studies should include toxicokinetics.

A recovery period should generally be included in study designs to determine the reversal or potential worsening of pharmacological/toxicological effects, and/or potential delayed toxic effects. For biopharmaceuticals that induce prolonged pharmacological/toxicological effects, recovery group animals should be monitored until reversibility is demonstrated. The duration of repeated dose studies should be based on the intended duration of clinical exposure and disease indication. This duration of animal dosing has generally been 1-3 months for most biotechnology-derived pharmaceuticals. For biopharmaceuticals intended for short-term use (e.g., ≤ to 7 days) and for acute life-threatening diseases, repeated dose studies up to 2 weeks duration have been considered adequate to support clinical studies as well as marketing authorization. For those biopharmaceuticals intended for chronic indications, studies of 6 months duration have generally been appropriate, although in some cases shorter or longer durations have supported marketing authorizations. For biopharmaceuticals intended for chronic use, the duration of long-term toxicity studies should be scientifically justified.

E. Immunotoxicity Studies (4.5)

One aspect of immunotoxicological evaluation includes assessment of potential immunogenicity (see section III.F (3.6)). Many biotechnology-derived pharmaceuticals are intended to stimulate or suppress the immune system and, therefore, may affect not only humoral but also cell-mediated immunity. Inflammatory reactions at the injection site may be indicative of a stimulatory response. It is important, however, to recognize that simple injection trauma and/or specific toxic effects caused by the formulation vehicle may also result in toxic changes at the injection site. In addition, the expression of surface antigens on target cells may be altered, which has implications for autoimmune potential. Immunotoxicological testing strategies may require screening studies followed by mechanistic studies to clarify such issues. Routine tiered testing approaches or standard testing batteries, however, are not recommended for biotechnology-derived pharmaceuticals.

F. Reproductive Performance and Developmental Toxicity Studies (4.6)

The need for reproductive/developmental toxicity studies is dependent upon the product, clinical indication and intended patient population (Note 2). The specific study design and dosing schedule may be modified based on issues related to species specificity, immunogenicity, biological activity, and/or a long elimination half-life. For example, concerns regarding potential developmental immunotoxicity, which may apply particularly to certain monoclonal antibodies with prolonged immunological effects, could be addressed in a study design modified to assess immune function of the neonate.

G. Genotoxicity Studies (4.7)

The range and type of genotoxicity studies routinely conducted for pharmaceuticals are not applicable to biotechnology-derived pharmaceuticals and therefore are not needed. Moreover, the administration of large quantities of peptides/proteins may yield uninterpretable results. It is not

expected that these substances would interact directly with DNA or other chromosomal material (Note 3).

Studies in available and relevant systems, including newly developed systems, should be performed in those cases where there is cause for concern about the product (e.g., because of the presence of an organic linker molecule in a conjugated protein product). The use of standard genotoxicity studies for assessing the genotoxic potential of process contaminants is not considered appropriate. If performed for this purpose, however, the rationale should be provided.

H. Carcinogenicity Studies (4.8)

Standard carcinogenicity bioassays are generally inappropriate for biotechnology-derived pharmaceuticals. However, product-specific assessment of carcinogenic potential may still be needed depending upon duration of clinical dosing, patient population, and/or biological activity of the product (e.g., growth factors, immunosuppressive agents, etc.). When there is a concern about carcinogenic potential, a variety of approaches may be considered to evaluate risk.

Products that may have the potential to support or induce proliferation of transformed cells and clonal expansion possibly leading to neoplasia should be evaluated with respect to receptor expression in various malignant and normal human cells that are potentially relevant to the patient population under study. The ability of the product to stimulate growth of normal or malignant cells expressing the receptor should be determined. When in vitro data give cause for concern about carcinogenic potential, further studies in relevant animal models may be needed. Incorporation of sensitive indices of cellular proliferation in long-term repeated dose toxicity studies may provide useful information.

In those cases where the product is biologically active and nonimmunogenic in rodents and other studies have not provided sufficient information to allow an assessment of carcinogenic potential, then the utility of a single rodent species should be considered. Careful consideration should be given to the selection of doses. The use of a combination of pharmacokinetic and pharmacodynamic endpoints with consideration of comparative receptor characteristics and intended human exposures represents the most scientifically based approach for defining the appropriate doses. The rationale for the selection of doses should be provided.

I. Local Tolerance Studies (4.9)

Local tolerance should be evaluated. The formulation intended for marketing should be tested; however, in certain justified cases, the testing of representative formulations may be acceptable. In some cases, the potential adverse effects of the product can be evaluated in single or repeated dose toxicity studies, thus obviating the need for separate local tolerance studies.

NOTES

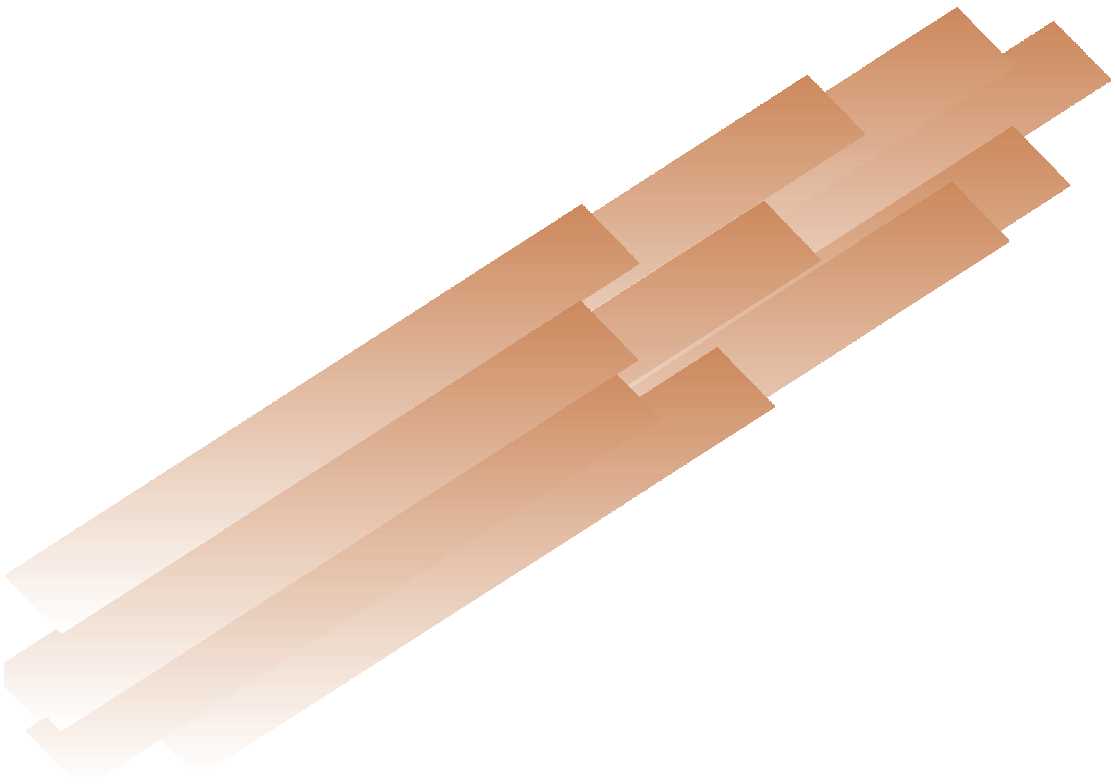
Note 1. Animal models of disease may be useful in defining toxicity endpoints, selection of clinical indications, and determination of appropriate formulations, route of administration, and treatment regimen. It should be noted that with these models of disease there is often a paucity of historical data for use as a reference when evaluating study results. Therefore, the collection of concurrent control and baseline data is critical to optimize study design.

Note 2. There may be extensive public information available regarding potential reproductive and/or developmental effects of a particular class of compounds (e.g., interferons) where the only relevant species is the nonhuman primate. In such cases, mechanistic studies indicating that similar effects are likely to be caused by a new but related molecule may obviate the need for formal reproductive/developmental toxicity studies. In each case, the scientific basis for assessing the potential for possible effects on reproduction/development should be provided.

Note 3. With some biopharmaceuticals, there is a potential concern about accumulation of spontaneously mutated cells (e.g., via facilitating a selective advantage of proliferation) leading to carcinogenicity. The standard battery of genotoxicity tests is not designed to detect these conditions. Alternative in vitro or in vivo models to address such concerns may have to be developed and evaluated.

Guidance for Industry

M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals



July 1997
ICH

Guidance for Industry

M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals

Additional copies are available from:
the Drug Information Branch (HFD-210),
Center for Drug Evaluation and Research (CDER),
5600 Fishers Lane, Rockville, MD 20857 (Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm
or
Office of Communication,
Training, and Manufacturers Assistance (HFM-40)
Center for Biologics Evaluation and Research (CBER)
1401 Rockville Pike, Rockville, MD 20852-1448,
http://www.fda.gov/cber/guidelines.htm
(Fax) 888-CBERFAX or 301-827-3844
(Voice Information) 800-835-4709 or 301-827-1800

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
July 1997
ICH

Table of Contents

GUIDANCE FOR INDUSTRY[1]

M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals

I. INTRODUCTION (1)

A. Objectives of the Guidance (1.1)

The purpose of this document is to recommend international standards for and to promote harmonization of the nonclinical safety studies needed to support human clinical trials of a given scope and duration.

Harmonization of the guidance for nonclinical safety studies will help to define the current recommendations and reduce the likelihood that substantial differences will exist between regions.

This guidance should facilitate the timely conduct of clinical trials and reduce the unnecessary use of animals and other resources. This should promote safe and ethical development and availability of new pharmaceuticals.

B. Background (1.2)

The recommendations for the extent of nonclinical safety studies to support the various stages of clinical development differ among the regions of Europe, the United States, and Japan. This raises the important question of whether there is scientific justification for these differences and whether it would be possible to develop a mutually acceptable guidance.

The present guidance represents the consensus that exists among the ICH regions regarding the scope and duration of nonclinical safety studies to support the conduct of human clinical trials for pharmaceuticals.

---

[1]This guidance was developed within the Expert Working Group (Multidisciplinary (Safety/Efficacy)) of the International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) and has been subject to consultation by the regulatory parties, in accordance with the ICH process. This document has been endorsed by the ICH Steering Committee at *Step 4* of the ICH process, July 1997. At *Step 4* of the process, the final draft is recommended for adoption to the regulatory bodies of the European Union, Japan and the United States. This guidance was published in the *Federal Register* on November 25, 1997 (62 FR 62922), and is applicable to drug and biological products. This guidance represents the Agency's current thinking on nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. An alternative approach may be used if such approach satisfies the requirements of the applicable statute, regulations, or both.

C. Scope of the Guidance (1.3)

The nonclinical safety study recommendations for the marketing approval of a pharmaceutical usually include single and repeated dose toxicity studies, reproduction toxicity studies, genotoxicity studies, local tolerance studies, and for drugs that have special cause for concern or are intended for a long duration of use, an assessment of carcinogenic potential. Other nonclinical studies include pharmacology studies for safety assessment (safety pharmacology) and pharmacokinetic (absorption, distribution, metabolism, and excretion (ADME)) studies. These types of studies and their relation to the conduct of human clinical trials are presented in this guidance.

This guidance applies to the situations usually encountered during the conventional development of pharmaceuticals and should be viewed as providing general guidance for drug development. Animal safety studies and human clinical trials should be planned and designed to represent an approach that is scientifically and ethically appropriate for the pharmaceutical under development.

There have been marked changes in the kinds of therapeutic agents being developed (e.g., biotechnology-derived products), and the existing paradigms for safety evaluation may not always be appropriate or relevant. The safety evaluation in such cases should be considered on a case-by-case basis as described in the ICH guidance "Safety Studies in Biotechnological Products" (Ref. 1). Similarly, pharmaceuticals under development for indications in life-threatening or serious diseases without current effective therapy may also warrant a case-by-case approach to both the toxicological evaluation and clinical development to optimize and expedite drug development. In these cases, particular studies may be abbreviated, deferred, or omitted.

D. General Principles (1.4)

The development of a pharmaceutical is a stepwise process involving an evaluation of both the animal and human safety information. The goals of the nonclinical safety evaluation include: A characterization of toxic effects with respect to target organs, dose dependence, relationship to exposure, and potential reversibility. This information is important for the estimation of an initial safe starting dose for the human trials and the identification of parameters for clinical monitoring for potential adverse effects. The nonclinical safety studies, although limited at the beginning of clinical development, should be adequate to characterize potential toxic effects under the conditions of the supported clinical trial.

Human clinical trials are conducted to demonstrate the efficacy and safety of a pharmaceutical, starting with a relatively low exposure in a small number of subjects. This is followed by clinical trials in which exposure usually increases by dose, duration, and/or size of the exposed patient population. Clinical trials are extended based on the

demonstration of adequate safety in the previous clinical trial(s) as well as additional nonclinical safety information that is available as the clinical trials proceed. Serious adverse clinical or nonclinical findings may influence the continuation of clinical trials and/or suggest the need for additional nonclinical studies and a reevaluation of previous clinical adverse events to resolve the issue.

Clinical trials are conducted in phases for which different terminology has been utilized in the various regions. This document uses the terminology as defined in the ICH guidance "General Considerations for Clinical Trials" (Ref. 2). Clinical trials may be grouped by their purpose and objectives. The first human exposure studies are generally single dose studies, followed by dose escalation and short-term repeated dose studies to evaluate pharmacokinetic parameters and tolerance (Phase I studies — Human Pharmacology studies). These studies are often conducted in healthy volunteers but may also include patients. The next phase of trials consists of exploratory efficacy and safety studies in patients (Phase II studies — Therapeutic Exploratory studies). This is followed by confirmatory clinical trials for efficacy and safety in patient populations (Phase III studies — Therapeutic Confirmatory studies).

II. SAFETY PHARMACOLOGY (2)

Safety pharmacology includes the assessment of effects on vital functions, such as cardiovascular, central nervous, and respiratory systems, and these should be evaluated prior to human exposure. These evaluations may be conducted as additions to toxicity studies or as separate studies.

III. TOXICOKINETIC AND PHARMACOKINETIC STUDIES (3)

Exposure data in animals should be evaluated prior to human clinical trials (Ref. 3). Further information on ADME in animals should be made available to compare human and animal metabolic pathways. Appropriate information should usually be available by the time the Phase I (Human Pharmacology) studies have been completed.

IV. SINGLE DOSE TOXICITY STUDIES (4)

The single dose (acute) toxicity for a pharmaceutical should be evaluated in two mammalian species prior to the first human exposure (Note 1). A dose escalation study is considered an acceptable alternative to the single dose design.

V. REPEATED DOSE TOXICITY STUDIES (5)

The recommended duration of the repeated dose toxicity studies is usually related to the duration, therapeutic indication, and scale of the proposed clinical trial. In principle, the duration of the animal toxicity studies conducted in two mammalian species (one nonrodent) should be equal to

or exceed the duration of the human clinical trials up to the maximum recommended duration of the repeated dose toxicity studies (Tables 1 and 2).

In certain circumstances, where significant therapeutic gain has been shown, trials may be extended beyond the duration of supportive repeated dose toxicity studies on a case-by-case basis.

A. Phase I and II Studies (5.1)

A repeated dose toxicity study in two species (one nonrodent) for a minimum duration of 2-4 weeks (Table 1) would support Phase I (Human Pharmacology) and Phase II (Therapeutic Exploratory) studies up to 2 weeks in duration. Beyond this, 1-, 3-, or 6-month toxicity studies would support these types of human clinical trials for up to 1, 3, or 6 months, respectively. Six-month rodent and chronic nonrodent studies (Ref. 11) would support clinical trials of longer duration than 6 months.

Table 1.—Duration of Repeated Dose Toxicity Studies to Support Phase I and II Trials in the EU and Phase I, II, and III Trials in the United States and Japan[1]

Duration of Clinical Trials	Minimum Duration of Repeated Dose Toxicity Studies	
	Rodents	Nonrodents
Single Dose	2-4 Weeks[2]	2 Weeks
Up to 2 Weeks	2-4 Weeks[2]	2 Weeks
Up to 1 Month	1 Month	1 Month
Up to 3 Months	3 Months	3 Months
Up to 6 Months	6 Months	6 Months[3]
> 6 Months	6 Months	Chronic[3]

[1] In Japan, if there are no Phase II clinical trials of equivalent duration to the planned Phase III trials, conduct of longer duration toxicity studies should be considered as given in Table 2.

[2] In the EU and the United States, 2-week studies are the minimum duration. In Japan, 2-week nonrodent and 4-week rodent studies are needed (Also see Note 2). In the United States, as an alternative to 2-week studies, single dose toxicity studies with extended examinations can support single dose human trials (Ref. 4).

[3] See Ref. 11. Data from 6 months of administration in nonrodents should be available before the initiation of clinical trials longer than 3 months. Alternatively, if applicable, data from a 9-month nonrodent study should be available before the treatment duration exceeds that which is supported by the available toxicity studies.

B. Phase III Studies (5.2)

For the Phase III (Therapeutic Confirmatory) studies, the recommendations for the United States and Japan are the same as those in Table 1. In the EU, a 1-month toxicity study in two species (one nonrodent) would support clinical trials of up to 2 weeks duration (Table 2). Three-month toxicity studies would support clinical trials for up to 1 month duration, while 6-month toxicity studies in rodents and 3-month studies in nonrodents would

support clinical trials of a duration up to 3 months. For longer term clinical trials, a 6-month study in rodents and a chronic study in nonrodents are recommended.

Table 2.—Duration of Repeated Dose Toxicity Studies to Support Phase III Trials in the EU and Marketing in All Regions[1]

Duration of Clinical Trials	Minimum Duration of Repeated Dose Toxicity Studies	
	Rodents	Nonrodents
Up to 2 Weeks	1 Month	1 Month
Up to 1 Month	3 Months	3 Months
Up to 3 Months	6 Months	3 Months
> 3 Months	6 Months	Chronic[2]

[1] The above table also reflects the marketing recommendations in the three regions except that a chronic nonrodent study is recommended for clinical use > 1 month.

[2] See Ref. 11.

VI. LOCAL TOLERANCE STUDIES (6)

Local tolerance should be studied in animals using routes relevant to the proposed clinical administration. The evaluation of local tolerance should be performed prior to human exposure. The assessment of local tolerance may be part of other toxicity studies.

VII. GENOTOXICITY STUDIES (7)

Prior to first human exposure, in vitro tests for the evaluation of mutations and chromosomal damage are generally needed. If an equivocal or positive finding occurs, additional testing should be performed (Ref. 5).

The standard battery of tests for genotoxicity (Ref. 6) should be completed prior to the initiation of Phase II studies.

VIII. CARCINOGENICITY STUDIES (8)

Completed carcinogenicity studies are not usually needed in advance of the conduct of clinical trials unless there is cause for concern. Conditions relevant for carcinogenicity testing are discussed in the ICH document (Ref. 7).

For pharmaceuticals developed to treat certain serious diseases, carcinogenicity testing, if needed, may be concluded postapproval.

IX. REPRODUCTION TOXICITY STUDIES (9)

Reproduction toxicity studies (Refs. 8 and 9) should be conducted as is appropriate for the population that is to be exposed.

A. Men (9.1)

Men may be included in Phase I and II trials prior to the conduct of the male fertility study since an evaluation of the male reproductive organs is performed in the repeated dose toxicity studies (Note 2).

A male fertility study should be completed prior to the initiation of Phase III trials (Refs. 8 and 9).

B. Women Not of Childbearing Potential (9.2)

Women not of childbearing potential (i.e., permanently sterilized, postmenopausal) may be included in clinical trials without reproduction toxicity studies provided the relevant repeated dose toxicity studies (which include an evaluation of the female reproductive organs) have been conducted.

C. Women of Childbearing Potential (9.3)

For women of childbearing potential there is a high level of concern for the unintentional exposure of an embryo/fetus before information is available concerning the potential benefits versus potential risks. There are currently regional differences in the timing of reproduction toxicity studies to support the inclusion of women of childbearing potential in clinical trials.

In Japan, assessment of female fertility and embryo-fetal development should be completed prior to the inclusion of women of childbearing potential using birth control in any type of clinical trial. In the EU, assessment of embryo-fetal development should be completed prior to Phase I trials in women of childbearing potential and female fertility studies prior to Phase III trials.

In the United States, women of childbearing potential may be included in early, carefully monitored studies without reproduction toxicity studies provided appropriate precautions are taken to minimize risk. These precautions include pregnancy testing (for example, based on the b-subunit of HCG), use of a highly effective method of birth control (Note 3), and entry after a confirmed menstrual period. Continued testing and monitoring during the trial should be sufficient to ensure compliance with the measures not to become pregnant during the period of drug exposure (which may exceed the length of study). To support this approach, informed consent should include any known pertinent information

related to reproductive toxicity, such as a general assessment of potential toxicity of pharmaceuticals with related structures or pharmacological effects. If no relevant information is available, the informed consent should clearly note the potential for risk.

In the United States, assessment of female fertility and embryo-fetal development should be completed before women of childbearing potential using birth control are enrolled in Phase III trials.

In the three regions, the pre- and postnatal development study should be submitted for marketing approval or earlier if there is cause for concern. For all regions, all female reproduction toxicity studies (Ref. 8) and the standard battery of genotoxicity tests (Ref. 6) should be completed prior to the inclusion, in any clinical trial, of women of childbearing potential not using highly effective birth control (Note 3) or whose pregnancy status is unknown.

D. Pregnant Women (9.4)

Prior to the inclusion of pregnant women in clinical trials, all the reproduction toxicity studies (Refs. 8 and 9) and the standard battery of genotoxicity tests (Ref. 6) should be conducted. In addition, safety data from previous human exposure are generally needed.

X. SUPPLEMENTARY STUDIES (10)

Additional nonclinical studies may be needed if previous nonclinical or clinical findings with the product or related products have indicated special safety concerns.

XI. CLINICAL TRIALS IN PEDIATRIC POPULATIONS (11)

When pediatric patients are included in clinical trials, safety data from previous adult human exposure would usually represent the most relevant information and should generally be available before pediatric clinical trials. The necessity for adult human data would be determined on a case-by-case basis.

In addition to appropriate repeated dose toxicity studies, all reproduction toxicity studies (Ref. 8) and the standard battery of genotoxicity tests (Ref. 6) should be available prior to the initiation of trials in pediatric populations. Juvenile animal studies should be considered on an individual basis when previous animal data and human safety data are insufficient.

The need for carcinogenicity testing should be addressed prior to long term exposure in pediatric clinical trials considering the length of treatment or cause for concern (Ref. 7).

XII. CONTINUING EFFORTS TO IMPROVE HARMONIZATION (12)

It is recognized that significant advances in harmonization of the timing of nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals have already been achieved and are detailed in this guidance. However, differences remain in a few areas. These include toxicity studies to support first entry into man and the recommendations for reproduction toxicity studies for women of childbearing potential. Regulators and industry will continue to consider these differences and work towards further improving the drug development process.

XIII. ENDNOTES (13)

Note 1 For the conduct of single dose toxicity studies, refer to the ICH-1 recommendations (Ref. 10) and the regional guidances.

Note 2 There are currently regional differences for the minimum duration of repeated dose toxicity studies; 2 weeks in the EU and the United States, and 2 weeks nonrodent and 4 weeks rodent in Japan. In Japan, unlike the EU and the United States, the male fertility study has usually been conducted prior to the inclusion of men in clinical trials. However, an assessment of male fertility by careful histopathological examination in the rodent 4-week repeated dose toxicity study has been found to be more sensitive in detecting effects on male reproductive organs than fertility studies (Ref. 9), and is now recommended to be performed prior to the first clinical trial in Japan. In the EU and the United States, 2-week repeated dose studies are considered adequate for an overall assessment of the potential toxicity of a drug to support clinical trials for a short duration.

Note 3 A highly effective method of birth control is defined as one that results in a low failure rate (i.e., less than 1 percent per year) when used consistently and correctly, such as implants, injectables, combined oral contraceptives, some intrauterine contraceptive devices (IUDs), sexual abstinence, or a vasectomized partner. For subjects using a hormonal contraceptive method, information regarding the product under evaluation and its potential effect on the contraceptive should be addressed.

XIV. REFERENCES (14)

1. ICH Topic S6 Document "Preclinical Testing of Biotechnology-Derived Pharmaceuticals."

2. ICH Topic E8 Document "General Considerations for Clinical Trials."

3. ICH Harmonised Tripartite Guideline (S3A) Note for "Toxicokinetics: The Assessment of Systemic Exposure in Toxicity Studies."

4. FDA, "Single Dose Acute Toxicity Testing for Pharmaceuticals; Revised Guidance," 61 FR 43934 to 43935, August 26, 1996.

5. ICH Harmonised Tripartite Guideline (S2A) "Guidance on Specific Aspects of Regulatory Genotoxicity Tests."

6. ICH Topic S2B document "Standard Battery of Genotoxicity Tests."

7. ICH Harmonised Tripartite Guideline (S1A) "Guideline on the Need for Carcinogenicity Studies for Pharmaceuticals."

8. ICH Harmonised Tripartite Guideline (S5A) "Detection of Toxicity to Reproduction for Medicinal Products."

9. ICH Harmonised Tripartite Guideline (S5B) "Toxicity to Male Fertility."

10. Arcy, P. F., and D. W. G. Harron, "Proceeding of The First International Conference on Harmonisation, Brussels 1991," Queen´s University of Belfast, pp 183-184 (1992).

11. ICH Topic S4 Document "Duration of Chronic Toxicity Testing in Animals (Rodent and Nonrodent Toxicity Testing)."

Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use

U. S. Department of Health and Human Services
Food and Drug Administration
Center for Biologics Evaluation and Research
February 28,1997

Date: February 27, 1997

From: Kathryn C. Zoon , Ph.D., Director
Center for Biologics Evaluation and Research
Food and Drug Administration

Subject: Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use

To: Manufacturers of Biological Products and Other Interested Persons

This Points to Consider (PTC) document has been developed for manufacturers of monoclonal antibody products for human use. These "Points" are not regulations nor are they guidelines, but represent the current thinking that the Center for Biologics Evaluation and Research (CBER) staff believe should be considered at this time. This 1997 PTC document supersedes the 1994 PTC document of the same title, announced in the Federal Register of August 3, 1994 (59 FR 39571).

It is our intention to continuously update and revise this document in order to improve its usefulness. We invite your review and comment on the "Points". Comments should be identified with the docket number 94D-0259. Two copies of any comments should be submitted except that individuals may submit one copy. All comments should be addressed to:

Dockets Management Branch (HFA-305)
Food and Drug Administration
12420 Parklawn Drive, Room 1-23
Rockville, MD 20857

_____-s-_____
Kathryn C. Zoon, Ph.D.

Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use

[Docket No. 94D-0259]

For further information regarding this document, contact:
Sharon Carayiannis
Center for Biologics Evaluation and Research (HFM-630)
Food and Drug Administration
1401 Rockville Pike, Suite 200N
Rockville, MD 20852-1448
301-594-3074

Submit written comments on this document to:

Dockets Management Branch (HFA-305)
Food and Drug Administration
12420 Parklawn Drive, Room 1-23
Rockville, MD 20857

Comments should be identified with the docket number found in the heading of this page.

Submit written requests for additional copies of this document or any other CBER guidance to:

Office of Communication, Training, and Manufacturers Assistance (HFM-40)
Food and Drug Administration
1401 Rockville Pike, Suite 200N
Rockville, MD 20852-1448

Send one self-addressed adhesive label to assist that office in processing your request.

These documents may also be obtained by mail by calling the CBER Voice Information System at 1-800-835-4709 or 301-827-1800, or by fax by calling the FAX Information System at 1-888-CBER-FAX or 301-827-3844.

Persons with access to the INTERNET may obtain these documents using, the World Wide Web (WWW), or bounce-back e-mail. For WWW access, connect to CBER at "http://www.fda.gov/cber/cberftp.html". To receive this document by bounce-back e-mail , send a message to "ptc__mab@A1.CBER.FDA.GOV".

TABLE OF CONTENTS

Note: Page numbering may vary for documents distributed electronically.

Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use February 1997

I. INTRODUCTION

A. BACKGROUND

Points to Consider documents provide a flexible approach in which FDA provides and updates its guidance on regulatory issues in many areas of drug development. Such documents are particularly useful in the rapidly evolving field of biotechnology-derived drugs and other biologics. The Center for Biologics Evaluation and Research (CBER) set out to revise the "Points to Consider (PTC) in the Manufacture and Testing of Monoclonal Antibody Products for Human Use" with several objectives. An important goal was to facilitate initial development of monoclonal antibodies for serious or life threatening indications. Additionally, it was felt that some of the information in the 1994 document required updating and streamlining. Finally, it was necessary to review this document for consistency with current CBER policy and with International Conference on Harmonisation (ICH) documents dealing with this category of products. This updated document supersedes the 1994 version, and is designed to assist sponsors and investigators regarding monoclonal antibody (mAb) product development, including information to submit when filing Investigational New Drug Applications ("INDs") and License Applications. Although this document does not create or confer any rights for or on any person and does not operate to bind FDA or the public, it does represent the agency's current thinking on monoclonal antibody products for human use.

For mAb, as for other biologics, certain regulations contained in 21 CFR Parts 200-299 and 600-680 apply and should be consulted. In common with the other PTC, the mAb PTC are not intended to be all-inclusive. They represent recommendations on how to conduct the clinical development of a product up to and after licensure, not checklists of items to be provided before or after phase 1 trials are initiated. Specific products which raise issues that are not considered in these "Points" will be evaluated on a case-by-case basis. The discussion on abbreviated product testing for feasibility trials in serious and immediately life-threatening conditions and on generic and modular virus clearance studies does not apply to human products made in human cell substrates. Consultation with CBER is strongly advised for sponsors considering the application of abbreviated testing policies to products that have the potential to be contaminated by human pathogens. For aspects of manufacturing and of the production facility that are not included in this discussion or in applicable regulations, sponsors should consult with the Office of Therapeutics Research and Review and the Office of Establishment Licensing and Product Surveillance respectively.

B. DEFINITIONS

For the purpose of this document, the terms "**antibody**" and **"monoclonal antibody"** (mAb) may be used interchangeably and refer to intact immunoglobulins produced by hybridomas, immunoconjugates and, as appropriate, immunoglobulin fragments and recombinant proteins derived from immunoglobulins, such as chimeric and humanized immunoglobulins, F(ab') and $F(ab')_2$ fragments, single-chain antibodies, recombinant immunoglobulin variable regions (Fvs) etc. Recommendations on the manufacture of recombinant products are contained in other PTC documents from CBER (1,2). Some of these recommendations pertaining to recombinant mAb produced in cell substrates other than hybridomas are reiterated in this document for convenience of consultation. This document applies to mAb used as therapeutic or *in vivo* diagnostic agents, as well as to **ancillary products**, i.e. mAb used in the manufacture of other products for *in vivo* use. The latter include mAb that are used alone or in

conjunction with devices, for example, for *ex vivo* purging of cells to remove immune or tumor cells, for *ex vivo* cell collection (e.g. hematopoietic stem cells), or for purification of other products intended for *in vivo* administration. Generally, these mAb should meet the same criteria for safety and freedom from adventitious agents as mAb intended for direct administration to patients. Likewise, reagents that are commonly used in conjunction with mAb for *ex vivo* manipulations of cellular products intended for *in vivo* administration (e.g. complement, DNAase) should meet the same safety standards as mAb intended for direct administration to patients. However, in such cases, some procedures for virus inactivation or removal may be performed on the downstream product rather than on the mAb or other reagent (see II.C.7). Complete information on products that will be used in conjunction with the mAb, such as rabbit complement or DNAase, should be submitted before clinical studies begin. This information can be submitted as a part of the original IND submission or in the form of a Master File.

As used in this document, "**cocktails**" are defined as two or more mAb administered at a fixed ratio. Relevant targets may include multiple antigens on infectious pathogens and multiple tumor-associated antigens. The rationale for combining the products should be clear and based on the clinical context or previous clinical experience with individual products. Lack of interference among the mAb in the combination should be shown and synergistic or additive effects should be characterized. Dose-ranging for each of the components is highly desirable. In some instances, dose-setting may be based on preclinical or clinical data that show the necessity or superiority of a particular dose and ratio of mAb in the combination.

As used in this document, "**panels**" are defined as sets of mAb directed against related antigens from which one or more members would be used for an individual patient based on target antigen characterization. Such panels could be submitted for approval in a single license application. Examples of panels might include anti-idiotype mAb for lymphoma and mAb directed against different bacterial or viral serotypes. Dose-ranging for each mAb would be necessary. During the phase 3 trials to establish efficacy of the entire panel, some clinical experience with each member of the panel should be obtained.

C. FILING INFORMATION

It is not necessary to have all of the information discussed in this document available in the initial IND submission. Rather, much of the information may be developed during clinical development, with guidance from CBER or other appropriate Centers by means of frequent dialogue. At pre-IND meetings, CBER staff may provide guidance in planning clinical development and establishing the format and content of initial IND submissions. Such meetings may be particularly useful when the product is a novel molecular entity or is produced by a novel process, and when drug development plans are unusually complex.

The manufacture of mAb that are produced and controlled by similar procedures in the same facility may in some cases be documented in a single Master File. This may be particularly helpful when data from generic or modular virus clearance studies are used for multiple antibodies that differ only in the variable (v) or complementarity-determining region (CDR) and when multiple antibodies are purified by identical procedures (see Section II.C.6).

See references 3 and 4 for information on filing biologics license applications. An Establishment License Application is no longer required for mAb intended for in vivo use (3).

II. PRODUCT MANUFACTURE AND TESTING

A. GENERAL PRINCIPLES AND DEFINITIONS

Traditionally, most mAb are produced by hybridoma cell lines through immortalization of antibody-producing cells by chemically-induced fusion with myeloma cells. In some cases, additional fusions with other lines have created "triomas" and "quadromas". We anticipate an increase in recombinant mAb (e.g. chimeric or humanized mAb, single-chain or dimeric Fvs, mAbs derived from phage display libraries etc.) and human mAb in the future. These may be produced in animal cell lines (e.g. CHO, SP2/0) transfected with recombinant DNA constructs, in human cells (e.g. immortalized lymphoid cells), in bacteria, yeast, insect cells etc. Novel methods of production for mAb or mAb-derived recombinant proteins may include insects, plants or transgenic animals,.
The principles reviewed in Sections II.B. 1 through 4 may be applied, in general to all hybridoma and heterohybridoma generated products, regardless of the species of origin. All steps in manufacturing of mAb to be used in trials intended to support licensure and of licensed mAb should comply with current Good Manufacturing Practices (cGMPs), as appropriate for the stage of product development.

While manufacturing details and safety issues may be different for different expression systems, some general principles can be applied. The establishment of a reliable and continuous source from which the antibody can be consistently produced is highly recommended (e. g. master cell banks for cell cultures, seed banks for transgenic plants, founder strains for transgenic animals). If transient expression systems are used, master vector seed stocks should be generated, and the genetic stability of the expression constructs used should be tested. Appropriate in-process testing which takes into consideration the specific safety concerns of the expression system used should be instituted. Sponsors are encouraged to consult the most recent available versions of the Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals, the Points to Consider in the Production and Testing of New Drugs and Biologicals produced by Recombinant DNA Technology or the Points to Consider in the Manufacture and Testing of Therapeutic Products for Human Use Derived From Transgenic Animals (1, 2, 5), the 1996 CBER/CDER Guidance Document on the Submission of Chemistry, Manufacturing and Controls Information for a Therapeutic Recombinant DNA-derived Product or a Monoclonal Antibody Product for In Vivo Use (4), as well as relevant International Conference on Harmonization (ICH) documents (e.g. 6, 7), if applicable to their expression systems. Sponsors considering novel expression systems not specifically covered by guidance documents are encouraged to consult with CBER.

B. MANUFACTURE AND CHARACTERIZATION OF MONOCLONAL ANTIBODIES

1. Cell lines

The following information should be provided in the IND or biologics license application:

a. Source, name, and characterization of the parent cell line, including any immunoglobulin heavy or light chains that it synthesizes and/or secretes, the fusion partner in the case of hybridomas, or the host cell line in the case of transfected cells producing recombinant mAb.

b. Species, animal strain, characterization, and tissue origin of the immune cell.

c. Description of immortalization procedures, if any, used in generating the cell line.

d. Identification and characterization of the immunogen. A complete biochemical characterization may not be possible or necessary in all cases. However, we recommend that as much information as possible be gathered on the nature and characteristics of the material used as an immunogen. Such data can be useful in choosing appropriate potency assays, as well as in evaluating potential for cross-reactivity and possible clinical usefulness. For example, a determinant which is not expressed on the surface of target cells bind necrotic cells better than intact cells.

e. Description of the immunization scheme. In the case of human mAb, any *in vitro* or *in vivo* immunization procedures should be described, as well as any relevant aspects of the subject's medical history.

f. Description of the screening procedure used. In the case of human mAb, steps performed in order to enrich antigen-specific human B cell populations should be described.

g. Description of the cell cloning procedures. If changes in cell culture process (e.g. cells adapted from serum-containing to serum-free medium) are shown not to affect product quality, it is not necessary to reclone the cells or rebank the MCB or WCB. In this context, product quality includes not just the identity, purity, potency and pharmacological characteristics of the purified product, but also its safety profile. For example, possible changes in types and/or titers of viruses detectable in the unpurified bulk material and the ability of the purification process to remove or inactivate them should be addressed).

h. For transfected animal or plant cell substrates, as well as for microbial cell substrates (bacteria, yeast), a detailed description of the vector(s) and final construct(s) generation, including whether or not extraneous amino acid sequences are introduced into the product as a result of subcloning, and description of transfection/transformation, screening and selection procedures (see refs. 1-6). Determination of cDNA sequence(s) of the predominant transcript(s) is acceptable as an indication of clonality of transfected cell lines.

i. For cell culture systems using autonomously replicating vectors (e.g. baculovirus or other transient expression systems) a detailed description of the vector system, construct generation, selection, vector banking procedures, and infection/transfection procedures should be provided (1-6).

j. For all cell substrates, description of the seed lot system for establishing and maintaining the master cell bank (MCB) and the working cell bank (WCB).

2. Production in cell culture

The following information should be provided in the IND or biologics license application:

a. A description of the culture procedures if production is entirely *in vitro* or if cells are passaged *in vitro* prior to mouse inoculation.

b. A description of the culture media used, including certification and testing. Serum additives used in hybridoma propagation should be free of contaminants and adventitious agents.

c. The steps taken to prevent or control contamination by viruses, bacteria, fungi, mycoplasma and transmissible spongiform encephalopaties (TSE) agents. These include, among other things, a description of the equipment, transfers, room classification, employee gowning procedures etc.

d. The acceptance criteria for cells or tissue culture supernatants intended for further manufacture.

3. Production in animals or plants

The following information should be provided in the IND or biologics license application:

a. A description of the cell line used as the inoculum (if any) should be provided (see 2.a. above).

b. Animal care should be in accordance with the NIH Guide for Care and Use of Laboratory Animals. For ascites production, the use of specific pathogen free (SPF) mice is recommended. To ensure manufacture of consistent, high quality ascites for production of mAb, an animal health monitoring program should be in place that encompasses quarantine procedures, sentinel animals, and an in-house health surveillance program (including screening for mycoplasma). Frequency of serological testing of sentinel mice should be established and is usually based on the incidence of virus contamination. Screening programs for known infectious agents should be updated to reflect advances in the knowledge of infectious diseases. Sponsors should be responsible for the adequacy of screening programs.

c. All protocols for ascites production should also incorporate information on: *i*) species, sex and age of animals used; *ii*) animal supplier; *iii*) volume of pristane; *iv*) volume and concentration of cell inoculum; *v*) timing of priming, inoculation, and ascites harvesting; *vi*) frequency and procedure for ascites harvesting; *vii*) definition of a batch; *viii*) animal bedding, food and water; *ix*) number of animals housed together; *x*) environmental conditions under which each procedure takes place and *xi*) number of times cells are passaged from one animal to another, if applicable.

d. For production in transgenic animals, the vectors, constructs and procedures used for gene transfer should be described. The genetic background and characterization of founder animals, the generation and selection of production herds and animal maintenance procedures should be described as well (see ref. 5 for details). Health monitoring programs for animal herds or colonies should be in place, including screening for zoonoses known to exist in captive animals of the relevant species in North America. Programs for screening and detection of known infectious agents should be tailored for the animal species and periodically updated to reflect advances in the knowledge of infectious diseases. Sponsors should be responsible for the adequacy of screening programs (see paragraph b above). When initially establishing transgenic animal strains, the following considerations should be kept in mind: *i*) non-transgenic animals to be used for breeding or gene transfer procedures should be obtained from closed herds or colonies that are serologically well characterized and as free as possible of pathogens of concern for the animal species or for humans; *ii*) the use of imported animals or first generation offspring from imported animals is discouraged and *iii*) animals from species in which TSE have been documented should be obtained from closed herds with documented absence of dementing illnesses and controlled food sources.

e. For production using autonomously replicating vectors (e.g. baculovirus) in live insect larvae, larvae maintenance procedures should be described in detail, including procedures used to control and monitor bioburden.

f. For production in plants (e.g. transgenic plants or autonomously replicating vectors using plants as bioreactors), early consultation with CBER staff is recommended.

4. Purification

Purification schemes for mAb should be described in the IND or biologics license application. We

recommend that mAb purification schemes incorporate the following characteristics:

a. Production techniques that will prevent the introduction of and eliminate contaminants, including animal proteins and materials, DNA, endotoxin, other pyrogens, culture media constituents, components that may leach from columns, and viruses.

b. Incorporation of one or more steps known to remove or inactivate retroviruses in excess of the endogenous particle load, whenever applicable (see Reference 8 and discussion of virus clearance studies in Section II.C.). As a general guidance, we recommend that each purification protocol include at least two orthogonal (i.e. based on different mechanisms) robust virus removal steps (see below). Including these steps would not obviate the need for virus clearance studies, except in the case of products intended for use in feasibility trials in serious or life-threatening conditions (see Section II.D.2.)

> *i.* **Robust virus removal/inactivation steps** are defined as those that have been shown to work well under a variety of conditions (e.g. pH or ionic strength of column buffers) with a variety of mAb. Robust steps include low pH, heat, solvent/detergent treatments, and filtration (see Table III). Sponsors have the option of providing adequate evidence indicating that a step different from these is robust, or is reliably effective for removal/inactivation under the conditions employed. An estimate of the efficiency of robust steps in removing virus may be demonstrated by: (a) cross-referencing master files or reliable scientific literature published in peer-reviewed journals or (b) generic or modular clearance studies (see Section II.C.6. for definition).

c. Demonstration of the ability of the purification scheme to remove adventitious agents and other contaminants, by means of a clearance study. For some contaminants, e.g. DNA, pristane or protein A, such a clearance study, if appropriately carried out, may be an acceptable alternative to routine testing for the contaminant. In the case of virus clearance studies, we recommend the use of several model viruses encompassing large and small particles, DNA and RNA genomes, as well as chemically sensitive and resistant lipid enveloped and non-enveloped strains. Human blood products should be avoided in production of other biologicals. When human blood products that may be contaminated with hepatitis C virus (HCV) or other infectious agents must be used in production (e.g. as media additives), such schemes should include viruses that are acceptable models for HCV, such as bovine viral diarrhea virus (BVDV) or Sindbis virus. Retrovirus clearance studies should be performed prior to phase 1 trials, except for products intended for use in the setting of serious or life-threatening conditions in feasibility trials (see Section II.D.). Clearance studies for other viruses and/or other contaminants should be carried out prior to production for phase 2/3 trials and may need to be repeated if the final manufacturing process has changed. ICH guidelines are currently being drafted to address in further detail the viral safety evaluation of biotechnology products derived from cell lines of human or animal origin.

d. Limits should be prospectively set on the number of times a purification component (e.g. a chromatography column) can be reused. Such limits should be based upon actual data obtained by monitoring the component s performance over time.

e. As a product is developed, retention samples from each production should be saved under appropriate conditions so that side-to-side comparisons may be made to determine product comparability (see Section II.E.).

f. A description of the purification room(s) design features, HVAC and other support systems, equipment, transfers and personnel should be provided. Emphasis should be placed on operational features that minimize the risk of contamination from the environment or cross-contamination from other products.

5. Characterization of purified unmodified mAb

Before a mAb is studied in humans, a precise and thorough characterization of antibody structural integrity, specificity, and potency should be conducted and described in the IND. The mAb should be as free as possible of non-Ig contaminants. A properly qualified in-house reference standard with known characteristics, specificity, and potency, and that is stored under appropriate conditions and periodically tested to ensure its integrity, should be used for lot-to-lot comparisons. Reference standards should be updated as a product evolves but should be finalized by the start of phase 3 trials. Appropriate standard operating procedures (SOPs) should be developed for qualification of a new reference standard.

a. Structural Integrity

A combination of SDS-PAGE, IEF, HPLC, mass spectrometry, or other appropriate physicochemical methods should be used to show that the purified antibody is not fragmented, aggregated, or otherwise modified (e.g. loss of carbohydrate side chains). Side-by-side comparisons of production lots to the in-house reference standard should be performed.

b. Specificity

Assays should provide evidence that the binding of the mAb to the target antigen is specific. Once the specificity of the antibody is characterized, it should be screened for cross-reactivity with human tissues (see Section III.A.). The following are some suggestions on the design of specificity studies:

i. Direct binding assays should include both positive and negative antibody and antigen controls. At least one isotype-matched, irrelevant (negative) control antibody should be tested. Negative antigen controls should include a chemically similar, antigenically unrelated compound, if available (e.g. similar chemical nature, size, charge, and charge density).

ii. Whenever possible, the protein, glycoprotein, glycolipid, or other molecule bearing the reactive epitope, should be biochemically defined, and the antigenic epitope, itself, determined. If the antigenic determinant is a carbohydrate, the sugar composition, linkage, and anomeric configuration should be established.

iii. If possible, fine specificity studies using antigenic preparations of defined structure (e.g. oligosaccharides or peptides) should be conducted to characterize antibody specificity by means of inhibition or other techniques. For complex biological mixtures, the lots of test antigen and/or inhibitors used for direct binding tests should be standardized. Inhibition of antibody binding by soluble antigen or other antibodies should be measured quantitatively.

iv. Once the specificity of an antibody has been determined, it is important to quantitate antibody binding activity by affinity, avidity, immunoreactivity, or combinations of these assays, as appropriate. A number of published methods are suitable for measurement of antibody binding activity (9, 10).

c. Potency Assays and Potency Specifications

Potency assays are used to characterize the product, to monitor lot-to-lot consistency, and to assure stability of the product. Potency may be measured by a binding assay, a serologic assay, activity in an

animal model, and/or a functional assay performed *in vitro* or *in vivo*. It is desirable that the assay(s) bear the closest possible relationship to the putative physiologic/pharmacologic activity of the product and be sufficiently sensitive to detect differences of potential clinical importance in the function of the product. In particular, when the performance of the antibody depends not only upon antigen binding but also on other critical functions, it is desirable that the potency assay(s) measure all such functions. Documentation of the potency assay's performance, including sensitivity, intra- and inter-assay variation and robustness, should be provided.

i. Antibody binding activity may be quantitated by ELISA, RIA, radioimmune precipitation, cytotoxicity, flow cytometry, or any other standard, appropriate method. Activity should be expressed as specific antigen-binding units per mg or g of antibody. Product should be compared to an in-house reference standard. Appropriate measurements of antibody affinity, if established, may be a useful adjunct to other assays. Parallel line bioassay or a similar, valid statistical procedure should be used in calculating potency.

ii. The potency of a mAb may also be tested by measurement of *in vivo* function in animal models, although such assays are often cumbersome and difficult to standardize and should not be the sole measure of potency.

iii. The permissible range of values in potency assays that reflects adequate biological activity of a product should be based on experience with a particular antibody. Ideally, potency assays should be correlated with *in vivo* activity in order to develop control tests which will ensure an effective product. This implies that multiple production lots should be used during the clinical development program and potency assay results should be correlated with clinical performance. When clinical performance is measured by *in vitro* tests used as surrogates of efficacy, such tests should be validated in a phase 3 clinical trial of appropriate design.

6. Anti-idiotype vaccines

The following issues should be addressed for anti-idiotype vaccines:

a. In the case of an anti-idiotype vaccine (Ab2 vaccine), the Ab2 immunogen should be characterized as to the Ab2 type, e.g. classical type (Ab2) or antigen mimic (Ab2) (11).

b. Ab2 vaccines should be shown to be reactive with the appropriate population of human Ab1 (antibody to nominal antigen) if such antibodies are available.

c. The Ab2 preparation should be studied for the appropriateness of response (to target antigen) in xenogeneic as well as syngeneic animals (12).

7. Monoclonal antibodies conjugated with toxins, drugs, radionuclides or other agents (immunoconjugates)

Immunoconjugates are typically produced by chemical processes using specific reagents to link the unconjugated antibody with a non-antibody agent. Alternatively, immunoconjugates can be obtained as chimeric recombinant proteins containing non-immunoglobulin and immunoglobulin sequences in the same polypeptide chain. In addition to previously discussed recommendations for unconjugated (naked) mAb, manufacturers of immunoconjugates should address the following:

a. Construction of the Immunoconjugate.

A full description of the reagents and the process used to construct an immunoconjugate should be submitted, including:

i. A description of components such as toxins, drugs, enzymes, and cytokines that are linked to the mAb, including: the source, structure, production, purity (including demonstration of freedom from adventitious agents), and characterization of all components (if components are purchased, a certificate of analysis should be supplied).

ii. A description of chemical components, such as linkers and chelating agents, that will be used in preparing the immunoconjugate. These should include documentation of the sources of reagents and method of preparation and determinations of residual impurities from synthesis or purification. Charts of the synthetic reaction pathways and any relevant published or in-house data concerning the toxicity of chemicals used in the production of an immunoconjugate should be provided.

iii. The average ratio of coupled material to antibody and the number of conjugated moieties per antibody should be determined as the first step in establishing lot release criteria for the final product and developing the relationship between immunoglobulin substitution number, potency, and stability.

iv. Products prepared using recombinant DNA technology (e.g., derived from transfected cell lines or microbial cell substrates, chimeric, reshaped, complementarity determining region [CDR] grafted, single chain Fv antibodies, and recombinant immunoconjugates) should follow recommendations discussed in references 1-7, as appropriate. The stability of recombinant immunoconjugates should be studied carefully, as such chimeric proteins may have altered conformational stability, solubility or tendency to aggregate compared to their component polypeptides in their native structures. Loss of specific immunoreactivity due to denaturation or formation of aggregates (e.g. diabodies formed by recombinant Fvs) may lead to altered pharmacokinetics and/or binding to non-target tissues.

b. Purity of the Immunoconjugate

i. Special care should be taken to ensure that the antibody preparations are as free as possible of extraneous immunoglobulin and non-immunoglobulin contaminants as such contaminants could react with nuclides, toxins or drugs during the construction of the immunoconjugate.

ii. The amount of free antibody and free components in the final product should be determined with limits set for each. Reactive intermediates should be inactivated or removed.

c. Immunoreactivity, Potency and Stability of the Immunoconjugate

Coupling of toxin or drug to an antibody may alter the activity of either component.

i. Immunoreactivity before and after coupling should be assessed using appropriate methodology (9, 10).

ii. Activity of the non-immunoglobulin component of immunoconjugates, should be assessed by a potency assay whenever appropriate (e.g., toxins, cytokines or enzymes, but not radio-

immunoconjugates intended for use in imaging)

iii. Limits on the percent change in immunoreactivity resulting from construction of the immunoconjugate should be established as part of product specifications.

iv. The immunoconjugate should be tested for stability *in vitro* by incubation in pooled human serum at 37° C under sterile conditions. Plasma may be used instead of serum, provided that the anticoagulant used does not affect the stability of the immunoconjugate (e.g., chelating agents may react with some radioisotopes, heparin may interact with basic proteins, etc.). Aliquots should be analyzed at timed intervals for the concentration of intact immunoconjugate and degradation products. The conditions under which product stability is evaluated and the positive and negative controls used should be fully described. Stability in human serum or plasma is not relevant for topically administered immunoconjugates which are demonstrated not to be absorbed into the bloodstream. It should be established whether or not such immunoconjugates or their components are detectable in plasma after topical administration, and whether or not they elicit an immune response.

d. Specific Issues Related to mAb Coupled to Radionuclides

The preparation of the radioimmunoconjugate should be performed in a standardized, well-controlled, and validated manner. Methods should be developed to estimate the percent radioactivity in each of the three species of concern: free isotope, conjugated mAb, and labeled, non-mAb substances.

i. It is recommended that the initial IND submission for a radiolabeled mAb contain analytical results from two to three radiolabeling runs that demonstrate the preparation of an immunoreactive, sterile, and pyrogen-free product. These radiolabeling runs should be performed by the same personnel who will radiolabel the mAb for the study, using the reagents that will be used for the study.

ii. Radiopharmaceutical grade isotopes should be used when preparing immunoconjugates. The sterility and pyrogen-free nature of each isotope should be documented by submission of a certificate of analysis and letters of cross-reference for manufacturing information.

iii. The concentrations of covalently-bound and free isotope in the final product as well as residual levels of labeling reagents and their decomposition products should be determined during the trial labeling runs.

iv. Quality control tests that will be performed before and/or after each patient administration should be described.

v. When appropriate, colloid formation by the radio-immunoconjugate should be determined and limits set for it.

C. QUALITY CONTROL AND PRODUCT TESTING

1. Cell line qualification

Qualification of the cell line for production of a mAb to be used as a biologic therapeutic should include screening the master cell bank (MCB) and the working cell bank (WCB), at least on a one-time basis, for endogenous and adventitious agents utilizing the tests outlined in Table I and described in the Points to

Consider in the Characterization of Cell Lines Used to Produce Biologicals (1). Because the WCB is derived from the MCB and propagated for only a few additional passages in tissue culture, abbreviated testing for detecting newly introduced contaminants is acceptable. Any virus contaminant should be quantified and, wherever possible, identified in order to establish the extent of virus clearance that the purification process should achieve (see also Section II.C.5.). The tropism of virus contaminants for human cells should be determined by appropriate infectivity assays. In the case of tissue culture or fermenter production, end-of-production cells (EPC) should be tested at least once to evaluate whether new contaminants are introduced or induced by the growth conditions. EPC should also be re-examined when there are changes in culture medium or in the scale of production. Cells at the limit of *in vitro* age used for production can be tested in lieu of EPC to allow for possible extensions in the length of time cells are kept in culture as manufacturing schemes are developed. The term EPC will be used throughout this text for ease of consultation. For cell lines which are known not to be susceptible to infection with mammalian viruses (e.g. plant cells, some insect cells), bacterial and fungal sterility testing, and in some cases testing for Mycoplasma or other mollicutes such as Spiroplasma, will be the most important concerns. Consultation with CBER is advised before using these cell substrates.

Table I
Cell Line Qualification

Tests	MCB	WCB	EPC
Sterility	+	+	+
Mycoplasma	+	+	+
Virus			
Adventitious	+	-	+
Species-specific*	+	-	-
Retrovirus	+	-	+
Authenticity	+	+	+

* Tests for rodent, primate, or human viruses (other than retroviruses), as appropriate
Retrovirus testing is not required for murine hybridomas. All other cell substrates should be tested as described in Section II.C.1.d.

a. Cell lines should be free from bacterial and fungal contamination as demonstrated by sterility testing. Recommended testing procedures for mycoplasma (cultivable and non-cultivable) are described in Ref. 1.

b. Screening for adventitious viruses (other than retroviruses) should include routine *in vivo* and *in vitro* tests (1).
c. Screening for species-specific viruses (other than retroviruses):

i. The mouse antibody production (MAP) test for murine cell lines (see Appendix II), the HAP test for hamster lines, and the RAP test for rat lines should be used. *In vivo* testing for lymphocytic choriomeningitis (LCM) virus, including non-lethal strains, is recommended. Testing of hamster cell lines should include minute virus of mice (see III.B.1.c.).

ii. Material that is contaminated with LCM, reovirus, Sendai virus, or Hantaan virus should not be used for mAb production.

iii. Cell lines from non-human primates should be screened for the following: herpes viruses (simiae and SA-8), cytomegalovirus (sCMV), encephalomyocarditis virus, simian hemorrhagic fever virus (SHF), varicella virus of simians (sVZV), adenovirus, SV-40, monkeypox, rubeola, and Ebola virus. Any other zoonotic agents suggested by the cell line derivation history should be screened for.

iv. Human cell lines should be screened for Epstein-Barr virus (EBV), cytomegalovirus (CMV), hepatitis B (HBV) and HCV, human herpes virus 6 (HHV-6), and any other viruses that are suggested by the medical history of the donor and type of tissue used to establish the original line. Cells from patients who are known to have developed Creutzfeld-Jakob disease (CJD) or other TSE, or from persons with two or more genetically related family members with CJD, should not be used.

v. Heterohybridomas using cells from 2 different species should be tested as appropriate for both species of origin.

vi. For cell lines of other species please consult with CBER.

d. Retrovirus testing of cell lines: Retrovirus contamination of cells from different species varies. The following should be considered when designing studies to detect retrovirus:

i. Murine cells used to produce monoclonal antibodies should be considered inherently capable of producing infectious murine retrovirus. The amount of retrovirus in the unprocessed bulk should be quantitated on a series of bulk harvests and shown to be consistent from lot to lot (1). Endogenous virus particle burden should be determined at the end of a typical fermentation, prior to purification, preferably by thin-section EM on material pelleted by ultracentrifugation. Particle burden determination is preferable to infectivity assays at this stage of production because it does not depend upon the susceptibility to infection of the cell lines used for virus amplification and it provides a "worst case scenario" of the level of viral contamination. Thin-section EM also allows morphological observation of viruses. Other, novel methods of equal or superior sensitivity and general applicability may be acceptable, if appropriately validated. Sufficient retrovirus removal by the purification scheme should be demonstrated (see also Section II.C.4.).

ii. Rat myeloma cell lines and hybridomas may not express retrovirus (13).The absence of detectable retrovirus, however, should be demonstrated by co-cultivation of the test article with a cell line(s) susceptible to a wide range of retroviruses combined with a sensitive detection assay, including examination of EPC and several production lots. If retrovirus is not detected by infectivity assays or electron microscopy, further clearance studies may not be needed. It is suggested that purification schemes for mAb produced by rat cell lines include one or more robust retrovirus inactivation or removal step.

iii. CHO cell lines express defective retrovirus particles (14). Whether hamster cell lines express infectious retroviruses has not been shown. Sponsors should demonstrate the lack of infectious hamster retroviruses by means of the most sensitive infectivity assays available. These include co-cultivation of the test article with a cell line(s) susceptible to a wide range of retroviruses combined with a sensitive detection assay. As a product moves into pivotal clinical trials, it may be necessary to make additional attempts to detect potential infectious virus by utilizing a wider range of indicator cells, including human cell lines (15, 16). Because of uncertainty about the validity of infectivity assays for hamster retroviruses, sufficient retrovirus particle removal by the purification scheme should be demonstrated (see also Section II.C.4-5). It is suggested that purification schemes for mAb produced by hamster cell lines include one or more robust retrovirus inactivation or removal step.

iv. Hybridomas or transfected clones produced from cells of non-human primate or human origin should be examined for the presence of retrovirus. Generic assays, such as transmission electron microscopy (TEM) or reverse transcriptase (RT) can be used to assess the presence of retrovirus. Other assays may be used, as long as they are appropriately validated. In addition, all primate cell lines should be screened for simian immunodeficiency virus (SIV), simian T lymphotropic virus (STLV), Foamy virus, human T lymphotropic virus (HTLV), and human immunodeficiency virus (HIV). Cell lines from non-human primates should be additionally tested for presence of simian retroviruses (SRV).

e. Each cell clone generated by stable transfection of widely used parental lines (e.g. CHO) should be considered as a new cell line from the standpoint of viral safety. Such clones may have significantly

different genetic characteristics compared to the parental line as a result of the transfection procedure itself, the clonal selection process and positional effects due to random integration of vector DNA into the cellular genome. Each clone should be screened for retroviruses and for adventitious viruses as described above and in Table I. Screening for species-specific viruses can be done once, on the MCB of the parental line.

f. Authenticity testing should confirm the cell line's species of origin, identity, and lack of cell-line cross-contamination.

2. Lot-to-lot quality control monitoring of unprocessed bulk lots and purified bulk lots, and final product specifications

Quality control monitoring should be performed on each lot of product, as defined in 21 CFR 600.3(x). Table II summarizes lot-to-lot product safety testing.

Table II
Lot-to-lot product safety tests

Tests	Unprocessed Bulk	Purified Product	Final Product
Sterility	+*	+	+
Mycoplasma	+	-	-
Virus			
Adventitious	+	-	-
Species-specific	+	-	-
Retrovirus	+	-	-
Polynucleotide	-	+	-
Endotoxin	-	-	+

*Bioburden testing with acceptable limits is sufficient at this stage.

In vitro testing with three indicator cell lines should be performed routinely for non-ascites material. *In vivo* testing is generally done once for non-ascites material but should be repeated when production methods change.

MAP, RAP or HAP testing for ascites only..

Quantitation of retrovirus (preferably by TEM) in the unprocessed bulk is important for murine hybridomas. For other hybridomas, generic assays for detection and quantitation of retrovirus, such as TEM coupled with appropriate co-cultivation assays are important if MCB or EPC are positive.

a. Unprocessed Bulk Lots

i. There should be set limits for bacterial bioburden in unprocessed bulk material. If bioburden testing of pooled ascites harvests shows the presence of viable contaminants, they should be quantified, and allowable limits for bacterial contamination should be set based on manufacturing experience. The identity of the bacterial species should be determined on a periodic basis and whenever the allowable limits for contamination have been exceeded. Filtration of ascites

harvests through a 0.45 m filter prior to storage is recommended (see also II.C.2.a.*ii*.).

ii. Tests for cultivable and non-cultivable mycoplasma should generally be performed on unprocessed bulk hybridoma supernatants, prior to any clarification by filtration (1). The filtration of unprocessed bulk ascites through a 0.45 m filter followed by storage at -60^{o} C prior to testing for mycoplasma is acceptable if samples of unfiltered material are retained for testing. If mycoplasma contamination of animals or unpurified bulk ascites or hybridoma supernatants is detected, these should not be used or processed further.

iii. In vitro virus testing with three indicator cell lines (e.g. Vero, MRC5, 3T3) should be performed routinely. *In vivo* testing is generally done once (as part of cell line qualification, Section II.C.1) but should be repeated when production methods change (1). Bioreactors containing hamster cells can become contaminated with minute virus of mice that may escape detection in routine *in vitro* assays. MAP testing or PCR testing for this virus appear to be more sensitive. In all cases, the frequency of monitoring should be specified in SOPs and justified based on actual experience when using continuous production in contrast to batch production. When contamination with a particular virus is encountered in a facility, consideration should be given to modifying the routine testing program in order to detect that virus.

iv. Species-specific virus testing should be performed (see Table II and II.C.1.c.).

v. Murine retrovirus contamination should be quantitated routinely for bulk ascites harvests. This may not be necessary if the sponsor can provide data showing that little variation exists in the concentration of mAb and retroviral load of their bulk ascites over several consecutive manufacturing runs, and the purification scheme used can remove substantially more than the highest load observed. If ascites production uses different groups of mice, periodic serologic monitoring for species-specific viruses should be performed on each group prior to their use for producing ascites. For tissue culture harvests, retrovirus contamination should be quantitated on three clinical grade production lots in order to establish the level of virus contamination for the specific cell line and manufacturing process (1) (see II.C.4). This quantitation of retrovirus should be done preferably by generic assays such as TEM or alternatively by sensitive infectivity assays (see also Section II.C.1.d.). Quantitation should be repeated when changes in tissue culture media, duration or scale of culture are made.

vi. For hybridomas of non-murine origin or other cell substrates, see section II.C.1.d. for appropriate assays to determine whether retrovirus is present. In those cases where MCB or EPC are positive for retrovirus, each lot of unpurified bulk should be examined for detection and quantitation of retrovirus by generic assays, such as TEM, coupled with appropriate co-cultivation assays.

b. Purified Bulk Lots (Drug Substance)

In addition to lot-to-lot safety testing summarized in Table II, routine testing on purified bulk lots of unmodified and modified mAb product should include the following determinations (for discussion of immunoconjugates see Section II.B.7.):

i. If a cell bank containing a known infectious agent is used, CBER staff should be consulted before proceeding with development, and clearance studies should be conducted to demonstrate the removal/inactivation of this agent by the mAb purification scheme. Testing for murine

retrovirus during clearance studies should employ infectivity assays which detect ecotropic recombinant murine retrovirus (ERV) and the polytropic or mink-cell focus-forming murine retrovirus (MCF). The infectivity assay should be comprised of an amplification period on a cell line sensitive to infection by these murine retroviruses (for example, Mus dunni cells, 17) coupled with an appropriate indicator assay (for example, PG4 S+L- assay for MCF virus, immunofluorescence assay with appropriate antibodies for detection of ERV). Assays which do not rely on infectivity, such as PCR-based assays, may be substituted, provided they have been validated for sensitivity and specificity, and that their results are correlated with those of infectivity assays. The first consecutive 3-5 lots of purified bulk should be tested to confirm that the contaminant was removed by the purification scheme. For cell lines containing viruses or virus-like particles, the absence of detectable virus should be confirmed for purified bulk by means of suitable methods having high specificity and sensitivity for the detection of the virus in question. For the purpose of marketing authorization, data from at least 3 lots of purified bulk manufactured at pilot-plant or commercial scale should be provided. However, for cell lines for which the endogenous particles have been extensively characterized, such as CHO cells, and adequate clearance has been demonstrated, it is not usually necessary to assay for the presence of non-infectious particles in purified bulk. Should a human infectious agent be identified in the cell bank, every lot of purified product should be tested and consultation with CBER staff is recommended before extensive product development.

ii. Chemical purity including the residual amounts of extraneous animal proteins, e.g., albumin, immunoglobulin or other contaminants in the final product. An SDS-PAGE analysis, under both reducing and non-reducing conditions, of increasing amounts of purified material should be provided. Generally, silver staining methods are more sensitive but less quantitative than Coomassie blue for SDS-PAGE.

iii. Molecular integrity, including the presence of aggregated, denatured or fragmented product.

iv. Immunoglobulin class or subclass, if used as a test of identity.

v. IEF pattern of the antibody (or its heavy and light chains) in each bulk lot with comparison to the in-house reference standard.

*vi***.** Sterility.

vii. Lot-to-lot testing for DNA content, prior to any excipient addition, is recommended as a way to monitor purification efficiency and reproducibility. DNA content in the final product should be as low as possible, as determined by a highly sensitive method. Low cell viability at harvest may contribute to high DNA levels in unprocessed bulk. It is suggested that, whenever possible, the final product contain no more than 100 pg cellular DNA per dose (18). It is suggested that a method with a sensitivity of 10 pg be used to determine DNA levels (2). An appropriately conducted clearance study for DNA removal may be an acceptable substitute for lot-to-lot testing.

viii. Tests for detection and quantitation of potential contaminants or additives (e.g., antibiotics, other media components, host cell proteins, chromatography reagents, preservatives, or components that may be leached from affinity chromatography columns such as protein A). Whenever possible, contaminants introduced by the recovery and purification process should be below detectable levels using a highly sensitive analytical method. However, for many of these potential contaminants, depending upon their potential for toxicity or immunogenicity, an

appropriately conducted clearance study may be an acceptable substitute for lot-to-lot testing. For products intended for marketing, at least 3 exhibit lots should be tested to confirm the removal of the contaminant(s) for which clearance studies have been conducted. Such clearance studies for product contaminants may have to be repeated when manufacturing schemes are changed. We recommend that antibiotics, particularly penicillin or other beta lactams, not be used. However, if they are used, their removal must be demonstrated by an adequate clearance study. The acceptability of trace contaminants that cannot be removed by standard methods should be discussed with CBER prior to the submission of an IND. Pristane, if used in the propagation of ascites fluid, should be shown to be undetectable by a sensitive test.

ix. A brief description of the formulation process and areas to be used for it should be provided. This description should incorporate information on the processing area, support systems, personnel and product transfers in sufficient detail to highlight the design or operational features utilized to minimize contamination or cross-contamination.

c. Final-Filled Product (Drug Product)

The following tests should be performed on the contents of final containers from each filling of product as defined in 21 CFR 600.3(y). In certain situations (e.g. user-radiolabeling), special approaches to final container testing may need to be developed on a case-by-case basis after discussion with CBER:

i. Protein quantity.

ii. Potency (21 CFR 610.10).

iii. Purity (21 CFR 610.13). Electrophoretic migration of the product in both the native and reduced states on polyacrylamide gels with comparison to the in-house reference standard can be used as a test for purity.

iv. Sterility (21 CFR 610.12).

v. A test for endotoxin. The Limulus Amebocyte Lysate (LAL) assay may be an acceptable equivalent method as allowed in 21 CFR 610.9 when validated by the rabbit pyrogen test as described in 21 CFR 610.13. A U.S. licensed test system should be used to perform the LAL assay. Comparative testing should be repeated when LAL lots made by a different manufacturer are used. Conditions necessary for comparative testing of the rabbit pyrogen and LAL assay procedures should be discussed with CBER on a case-by-case basis. See also "Guidelines on Validation of the LAL Test as an End-Product Endotoxin Test for Human and Animal Parenteral Drugs, Biological Products, and Medical Devices", December , 1987 (19).

vi. An appropriate identity test (21 CFR 610.14).

vii. Moisture (21 CFR 610.13) testing, when appropriate.

viii. Preservative (21 CFR 610.15) testing, when appropriate.

ix. Excipients, when appropriate

x. pH, when appropriate.

xi. The areas used for final fill of drug products should follow the recommendations provided in the Guidelines on Sterile Drug Products Produced by Aseptic Processing (20). For information on validation of equipment used in aseptic processing, refer to the Guidelines for Submitting Documentation for Sterilization Process Validation in Applications for Human and Veterinary Drug Products (21).

3. Stability of product

Product stability should meet the demands imposed by the clinical protocol. Accelerated stability testing data may be supportive but do not substitute for real-time data for product approval and labeling.

a. A stability testing program should be developed that includes tests for physico-chemical integrity (e.g., fragmentation or aggregation), potency, sterility, and, as appropriate, moisture, pH and preservative stability, at regular intervals throughout the dating period. See also "Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics" (22) and the relevant ICH document (7). For products that are in clinical trials, significant changes that occur during storage should be reported to CBER. For license applications, and preferably prior to each stage of development, stability tests that support the proposed dating period should be performed on the final-filled product, using the container and closure configuration intended for distribution. For biologics license applications, storage of production intermediates (e.g. unpurified bulk, purified bulk) also should be supported by stability testing.

b. Stability tests for assuring biological activity (e.g. quantitative *in vitro* potency assays) should include a manufacturer's in-house reference standard. Whenever possible, a single lot of test antigen (e.g. purified antigen, cells or tissue) should be used throughout the study. A quantitative potency assay(s) should be used to permit a meaningful comparison of activities.

c. Accelerated stability studies, i.e. stability testing after storage at temperatures exceeding routine storage temperatures, may help to identify and establish which tests are stability-indicating. Specific parameters that indicate stability should be monitored by trend analysis on every lot on the stability testing program.

4. General considerations on quantitation and removal of a retrovirus contaminant

The amount of any retrovirus contaminant in the unprocessed bulk product should be estimated. If a retrovirus is detected, any possible or suspected tropism for human cells should be explored (e.g. by co-cultivation assays). This should be followed by demonstration of removal of the contaminant, using the contaminant itself or a representative analogue of the known contaminant (e.g. a model retrovirus) in a model purification system (1).

Unprocessed bulk supernatant concentrates or ascites should be assayed prior to any manipulation other than clarification by low speed centrifugation, unless it can be shown that virus testing would be made more sensitive by initial partial processing. We recommend that retroviruses contaminating ascites or supernatants produced by rodent cell lines be quantitated by TEM of concentrated samples of supernatant. If TEM results are negative, it should be assumed that the titer of retrovirus is equivalent to the lowest limit of detection ($1x10^6$/ml). Ascites and supernatants produced by non-rodent or hybrid cell substrates should be assayed as described (see Section II.C.2.b.*vi.*). Demonstration that retroviruses are removed or inactivated by the purification scheme should occur prior to phase 1 studies, except for mAb intended for use in feasibility trials in serious or life-threatening conditions (see Section II.D.). References 1, 2 and 8,

and paragraphs 5 and 6 below, also discuss the design of studies to demonstrate retrovirus removal. The goal of such studies is to demonstrate that the purification process is able to remove substantially more virus than is estimated to be present in a single dose equivalent of starting material.

5. General considerations on the design and interpretation of virus clearance studies

These considerations are provided here solely for guidance purposes in the design of virus clearance (i.e., removal/inactivation) studies.

a. General Experimental Design

The objective of a virus clearance study is to provide a quantitative estimate of the level of virus reduction provided by the removal/inactivation procedures. Thus, the study should be designed, conducted and analyzed in a manner that will provide accurate information to reliably assess the ability of these procedures to remove/inactivate viruses. The study, including virus infectivity assays, should be designed according to good scientific practice to yield data with accuracy and precision that are amenable to statistical analysis. Each clearance study should include an appropriate control experiment performed in parallel to the experimental condition to assess virus inactivation caused by experimental manipulation (dilution, concentration, filtration, storage etc.). Any observed difference should be used to adjust the virus reduction/inactivation values for each removal/inactivation procedure. It is preferable that the study be conducted under current Good Laboratory Practices (21 CFR, Part 58).

b. Statistics

Virus clearance studies should be supported by appropriate statistical analysis demonstrating that the study is valid and reliable. Virus infectivity assays used to quantitate the virus titer should be sensitive, reproducible and conducted with sufficient replicates to demonstrate statistical accuracy. Sufficient sample volumes should be tested to ensure that there is a high probability of detecting virus in the sample if present. The experimental variability present in virus titration should be determined in virus infectivity assays and reported using confidence intervals. Within assay variability should be calculated using conventional means (standard deviation, standard error of the mean, etc.). Between assay variation can be monitored by the inclusion of an appropriate virus reference preparation run in parallel with unknown samples. The 95% confidence intervals for within assay variation should be established and reported for each virus infectivity assay, and should be in the order of ± 0.5 $\log_{10}$ of the mean. The quantitative estimate of virus reduction for each procedure should be reported as the reduction factor. This is defined as the $\log_{10}$ of the ratio of the virus load in the starting material to the virus load in the post reduction material ($\log_{10}$ (volume x virus titer of starting material / volume x virus titer of post-reduction material)). In general, this virus reduction factor should be based upon the amount of virus detected in the spiked starting material. The 95% confidence interval for each reduction factor is calculated from the 95% confidence interval of the virus infectivity assays at the beginning (+s) and end (+a) of each procedure using the formula ± $\sqrt{a^2+b^2}$. Such confidence intervals should be calculated whenever possible in studies of clearance using relevant or specific model viruses (with respect to viruses actually detected in the cell line as opposed to non-specific model viruses). The reduction factors for each procedure are summed to calculate an overall reduction factor for the entire process. Additivity of reduction factors assumes that different steps are independent, i.e. have different mechanisms of action so that virus forms (mutant, aggregates etc.) escaping one step do not have a higher likelihood of also escaping the other. Reduction factors from two steps with the same mechanism of action (e.g. two incubations at low pH) are not necessarily additive. Due to the high intrinsic variability of infectivity assays, reduction factors of one log or less are considered negligible and are not included in the overall reduction factor.

6. Generic or modular virus clearance studies

a. A generic clearance study is one in which virus removal and inactivation is demonstrated for several steps in the purification process of a model antibody. These data may then be extrapolated to other antibodies following the same purification and virus removal/inactivation scheme as the model antibody.

b. A modular clearance study is one that demonstrates virus removal or inactivation of individual steps during the purification process (column chromatography, filtration, pasteurization, solvent/detergent, low pH, etc). Each module in the purification scheme may be studied independently of the other modules. Different model mAb may be used to demonstrate viral clearance in different modules, if necessary. If the purification process of a product mAb differs at any of the virus removal or inactivation modules from the model mAb, this module must be studied independently from the model. The other, identical modules in the procedure may be extrapolated to the product mAb.

c. Applicability: For monoclonal antibodies manufactured at one site, generic or modular clearance data may be extrapolated to other monoclonal antibodies of the same species and H and L chain class and, generally, subclass, derived from the same source (e.g. ascites or tissue culture) and cell substrate. Generic and/or modular clearance will apply only when the mAb have similar biochemical properties and are purified by identical methods. Particular attention should be paid to column elution buffer conditions, including pH and ionic strength, the sequence of columns, protein concentration, dwell times, flow rate, pressure, temperature, and potential problems associated with scale-up, at all steps of virus inactivation and removal. In some cases, sponsors may demonstrate virus removal/inactivation for a particular module at two different values of a given parameter (e.g. ionic strength, dwell time, temperature) and use any values of that parameter falling within this range. In order to apply generic or modular clearance algorithms to a specific product, it is necessary to determine the virus load in the unpurified bulk for each specific product using at least 3 lots (see Section II.C.2.). This information is then used to determine whether the total reduction/inactivation factor provided by the purification procedure ensures that substantially more virus is removed or inactivated than the estimated unpurified bulk titer. For example, generic clearance might apply to a series of murine mAb of different specificities but of the same H and L chain isotypes that are purified in an identical manner, or to a series of humanized mAb of the same H and L chain isotype but with different CDRs, that are purified by identical methods. The concepts of generic and modular virus clearance studies do not apply to products of entirely human origin or to products that have the potential to be contaminated by human pathogens. Consultation with CBER is advised before applying generic or modular virus clearance studies.

d. Virus Clearance Master Files: A sponsor purifying mAb for a variety of applications or a manufacturer producing mAb for a variety of sponsors may submit a Master File containing the data demonstrating virus removal or inactivation for different purification schemes. This Master File may then be cross-referenced in INDs/IDEs or license applications using mAb purified by these schemes if applicability criteria are met.

7. Product testing requirements for mAb used as ancillary products

Testing requirements for mAb used as ancillary products (see I.B. for definition) vary depending on the clinical indication and the stage of product development. From a safety standpoint, they should be characterized in the same way as products intended for *in vivo* administration. However, production steps which follow purification of the mAb can be used as part of the virus removal/inactivation scheme. These include, for example, conjugation of mAb to solid phases for affinity purification, sanitation of affinity columns etc. The concepts of generic and modular clearance studies apply to ancillary products as well. Purity of the ancillary product may not be as critical as for products intended for *in vivo* administrations, provided that the performance of the ancillary product in the production of the final product is acceptable

and reproducible, limits are set for impurities and the nature of impurities is known. Leaching of mAb or impurities from the mAb preparations into the final product should be taken into consideration in testing the final product, as appropriate. Removal of mAb or impurities from the mAb preparation may be demonstrated by means of a clearance study. Labeling for the final product may need to carry precautionary statements about potentially toxic or immunogenic residual impurities.

D. PRODUCT SAFETY TESTING FOR FEASIBILITY CLINICAL TRIALS IN SERIOUS OR IMMEDIATELY LIFE-THREATENING CONDITIONS

1. General considerations

The extent of product testing necessary before a particular clinical trial is initiated depends on the source and nature of the product, the stage of product development and the clinical indication. Abbreviated testing needs described in this Section (II.D.) apply to feasibility clinical trials in serious or immediately life-threatening conditions for which no effective alternative treatment exists. Abbreviated testing does not apply to human products made in human cell substrates, but may apply to recombinant products made from transfected human genes, depending on the cell substrate. Consultation with CBER is strongly advised for sponsors considering the application of abbreviated testing in this setting to products that have the potential to be contaminated by human pathogens. For the purpose of this document, the following definitions should be considered:

> *i*) **Feasibility clinical trials.** These are pilot studies whose objectives include, among others, early characterization of safety and initial proof of concept in a specific patient population. These trials are limited in scope, and are generally conducted in a single center, with a small number of patients (e.g. 5-20). These trials cannot be used by themselves to support licensure of a product. Studies conducted in normal volunteers are not included in this definition.
>
> *ii)* **An immediately life-threatening condition** is "a stage of a disease in which there is a reasonable likelihood that death will occur within a matter of months or in which premature death is likely without early treatment" (21 CFR 312.34).

See 21 CFR 312.34 and the Federal Register vol. 52, No. 99 (May 1977) for a discussion of serious or life-threatening conditions. Application of abbreviated testing requirements to serious conditions which are not immediately life-threatening as defined in 21 CFR 312.34 will depend upon an assessment of the potential risks and benefits to the patient(s). Factors that should be considered in such a risk-benefit analysis include, among others: *i*) the nature and manufacture of the product; *ii*) the nature and severity or stage of the disease; *iii*) the anticipated effect(s) of product administration (e.g., diagnosis, palliation or cure); *iv*) the availability of comparable or satisfactory alternate treatments *v*) characteristics of the patient population (e.g. age, response to previous therapy); *vi*) the number of patients involved and *vii*) the design of the clinical trial (e.g. patient follow-up, safety monitoring etc.). Pre-IND consultation with CBER staff is strongly recommended for sponsors planning to use the abbreviated testing described below for serious but not immediately life-threatening diseases. The guiding principle for these trials is that sufficient information should be provided before testing in human subjects to assure that patients and their contacts will not be put at unacceptable risk. Informed consent issues for these trials are discussed in Section IV.A.1.e. The limited testing described below should not be used to support development beyond the stage of feasibility trials. Therefore, sponsors are encouraged to plan for additional testing and characterization as described in Section II.C. when they intend to pursue advanced clinical development and seek licensure. In designing the purification process, it is advisable to include at least two orthogonal robust virus inactivation/removal steps. This would further reduce the testing necessary to begin initial clinical trials.

2. Product safety data needed before the initiation of feasibility trials in serious or immediately life-threatening conditions

a. Sterility (bacteria and fungi) testing should be performed on the final product. Three vials of final product should be tested. Routine methods in use in the sponsor's hospital accredited clinical diagnostic laboratory can be used for these tests. Mycoplasma and endotoxin testing are strongly encouraged.

b. *In vitro* and *in vivo* testing for adventitious viruses : If the unpurified bulk product is free of adventitious viruses by *in vitro* and *in vivo* tests or if the purification scheme does include at least two orthogonal robust steps (see Table III), these tests are not required.
Table III shows ranges of retrovirus removal (expressed in decimal logs) that might be expected with various robust inactivation/removal steps and is presented to aid manufacturers in the design of mAb purification schemes.

Table III

Inactivation Step	Reported log virus removal
pH ≤ 3.9	**3-4**
heat	**4**
solvent/detergent	**5**
filtration (15-40 nm)	**4-8**

If the mAb is produced as an ascites fluid and the mice used have been MAP tested and found free of species-specific viruses, adventitious virus testing of the final product is not necessary. When testing is necessary, we recommend that a minimum amount of product equal to 3 maximum human doses be used for these tests. For the purpose of these trials, MAP testing can be limited to known human pathogens: Hantaan, LCM, Reovirus and Sendai virus. For mAb produced in primate cell lines or in non-murine cell lines whose potential for contamination by human pathogenic viruses in unknown, the cell lines or the EPC or the unpurified bulk product should be tested as described in Section II.C.1. c and d.

c. Murine retrovirus testing of the final product is needed on the final-filled product only when the antibody is produced in a murine cell substrate and the purification scheme does not include at least two robust orthogonal virus inactivation/removal steps (see Table III). We recommend that a minimum amount of product equal to 3 maximum human doses be used for these tests. Testing of the final product, when necessary, should be done by a highly sensitive infectivity assay, such as amplification in Mus dunni cells followed by detection in PG4 cells or by other sensitive means as outlined in II.C.2.b.*vi.*

d. MAb used as ancillary products (see I.B. for definition) in feasibility clinical trials in serious or life-threatening conditions. Two cases can be distinguished:

> *i.* When the final product can be tested and the results are available prior to administration (e.g. purified recombinant proteins or cells that can be stored frozen), safety testing can be carried out on the final product itself, as determined during review of the final product.

ii. When the final product is administered prior to or without any safety testing and/or processing, testing of the mAb should be performed as described above in paragraphs D.2. a. through c. In this case, amounts of mAb comparable to those used in one run of final product purification should be used.

The same safety considerations apply to complement, DNAase and other biological reagents used as ancillary products for cell depletion in conjunction with mAb.

E. ISSUES RELATED TO MANUFACTURING CHANGES (DEMONSTRATION OF PRODUCT COMPARABILITY)

1. General

Changes in the product manufacturing scheme frequently occur during clinical development of mAb. Sponsors should develop a plan for demonstrating that the products made by the old and new schemes are comparable, particularly when preclinical or clinical data developed prior to the production changes will be used to support further clinical trials and/or marketing applications (23). Similar considerations apply in the case of significant scale up in the manufacturing process (with or without modification of the general manufacturing scheme) implemented during or after completion of phase 3 trials.

When changes in manufacturing occur during early clinical development, plans for evaluation of product comparability should be incorporated into product development strategies. Such plans should be discussed with CBER and, when appropriate, submitted to CBER for review (see ref. 23).

In-process specifications may be affected by manufacturing changes or process scale-up and appropriate revisions should be undertaken. Similarly, process validations (e.g., virus clearance studies, removal of contaminants or leachables) for all affected steps should be repeated after any significant manufacturing change. In the case of process scale-up, it is recommended that, whenever possible, a column geometry and a ratio between sample volume and bed volume as close as possible to that of the original process be maintained. This is particularly important for those steps (e.g., size-exclusion chromatography) where these parameters are critical to the chromatographic process. Ref. 23 should be consulted for more details on demonstrating product comparability.

2. In vitro evidence of product comparability

In general, when a product is obtained by a modified or scaled-up manufacturing scheme, the results of a rigorous physico-chemical characterization and *in vitro* functional comparison (see Product Manufacture and Testing, Section II.B.) will dictate whether additional data (e.g., pre-clinical and/or clinical data) will be needed. A protocol for demonstrating physico-chemical, immunochemical and biological comparability of two products should prospectively define acceptable variation in the results of individual assays and acceptance criteria for product comparability. For quantitative assays (specific biological assays) accurate estimates of inter and intra-assay variations should be provided. Assays with high intrinsic variability are poorly suited for the evaluation of product comparability. Comparisons should test a number of separate product lots in parallel in order to demonstrate the reproducibility of the new manufacturing scheme. An *in vitro* biochemical characterization of mAb comparability should include a side by side comparison of the two products by a number of different techniques. Properly stored retention samples from previous lots should be used for such side by side comparison. A list of techniques could include SDS-PAGE under reducing and non-reducing conditions, Western blot, size-exclusion analytical chromatography, reverse phase high performance liquid analytical chromatography, isoelectrofocusing, mass spectrometry, an

analysis of glycosylation including carbohydrate content and composition, peptide mapping or other appropriate tests. *In vitro* functional comparison should include assays aimed at the characterization of the biological function of the antibody (e.g., binding, cytotoxicity, epitope modulation, etc.). Whenever possible, a comparison of the affinity constants of the two products is highly recommended.

3. Animal studies

Depending on the quality of the data and the type of *in vitro* assays, the nature of the manufacturing change and the types of product differences observed or anticipated, a program of comparative testing (pharmacokinetics, etc.) in appropriate animal models may be considered in lieu of human clinical data when biochemical testing shows differences or cannot exclude significant differences in two products. In some cases, pharmacokinetic studies are complementary to *in vitro* studies. Pharmacokinetic studies in animals may be informative, even in the absence of the target antigen, depending upon the question to be addressed and the expected contribution of antigen binding to the biodistribution of specific mAb in humans. The extent of animal toxicity testing that may be needed to assess comparability will depend upon the safety profile of each specific product, the magnitude of the changes in manufacturing, the presence or absence of detectable differences in purity, structure or *in vitro* activity. Sponsors are encouraged to discuss plans for comparative testing of the two products in animals with CBER or to submit proposal for such testing to CBER for review and comment. The proposed program should be appropriate in view of biochemical data and include statistical considerations.

4. Clinical studies to support manufacturing changes

Comparative clinical evaluation of the products produced by different or scaled-up manufacturing schemes may be needed in certain situations:

a. Product activity cannot be adequately characterized by analytical testing.

b. Biochemical or biological testing show differences in the products.

c. Animal testing reveals significant pharmacokinetic or other differences in the products.

d. The formulation of the product has been changed in a way that can affect its bioavailability. The latter changes generally dictate a need for clinical pharmacokinetic studies.

Pharmacokinetic, safety and/or efficacy data may be required depending upon the nature and magnitude of the observed changes in the biochemical and or biological properties of the product.

Additional information on product comparability testing can be found in ref. 23.

III. PRECLINICAL STUDIES

A. TESTING CROSS-REACTIVITY OF MAB

When the same or related antigenic determinant is expressed on human cells or tissues other than the intended target tissue, binding of the antibody to this tissue may be observed. Non-target tissue binding may have serious consequences, particularly when pharmacologically active antibodies or cytotoxic immunoconjugates are used. Accordingly, cross-reactivity studies with human tissues (or cells if

applicable) should always be conducted prior to phase 1 to search for cross-reactions or non-target tissue binding. In the special case of bispecific antibodies, each parent antibody should be evaluated individually, in addition to testing of the bispecific product.

1. *In vitro* testing for cross-reactivity

Human cells or tissues are presently surveyed immunocytochemically or immunohistochemically. Appropriate newer technologies should be employed as they become available and validated.

a. Reactivity of the antibody or immunoconjugate should be determined with the quick-frozen adult tissues listed in Appendix I. Surgical samples are preferred. Post-mortem samples are acceptable with adequate tissue preservation. Tissues from at least three unrelated human donors should be evaluated in order to screen for polymorphism. The effect of fixatives on tissues that are known to be positive should be evaluated to ensure that the target antigen is preserved during tissue processing.

b. In special situations it may be appropriate to assay cross-reactivity on representative cultured cell lines, stem cells, and embryonic/fetal tissue.

c. Several concentrations of the product should be tested. The ability to detect cross-reactions may depend on antibody concentration. Antibody affinities as well as expected achievable peak plasma concentrations should be considered when choosing the proper concentrations for tissue binding studies. An "ideal" concentration for these studies may be the lowest mAb concentration that produces maximum (plateau) binding to the target antigen. An attempt should also be made to compare the ratio of specific binding to target tissue to specific binding to cross-reactive tissue. Because non-specific as well as Fc-mediated binding may be observed, it should be distinguished from specific cross-reactions using inhibition assays with purified antigen, when available.

d. Positive and negative controls are essential for interpreting study results. Controls confirm the acceptable condition of the tissues and adequacy of the assay. Anti-transferrin receptor mAb may be a useful positive control, since transferrin receptor is a common and abundant molecule on the surface of growing normal and tumor cells.

e. If a conjugated, chemically modified antibody or antibody fragment is to be used clinically, it should be tested in that form if at all feasible. The substitution of antibodies of similar specificity for cross-reactivity testing is discouraged.

f. When cross-reactions are encountered and there is a reason to suspect genetic polymorphism of the target antigen, studies should be expanded to a larger panel of tissues to better characterize this polymorphism.

g. A comparison of *in vitro* cross-reactivity in tissues from different species is important in determining the most relevant animal for subsequent toxicology studies.

2. *In vivo* testing for cross-reactivity

Cross-reactivity of a monoclonal antibody with non-target human tissues should dictate a comprehensive *in vivo* investigation in animals, when appropriate models are available. This finding, particularly with cytolytic immunoconjugates or antibodies with ADCC activity, generally indicates the desirability of more extensive preclinical testing, including studies in more than one animal species over a range of doses and

repeat dose animal studies. Localization to non-target tissues should be kept in mind when designing clinical trials.

B. PRECLINICAL PHARMACOLOGY AND TOXICITY TESTING

1. General considerations

a. Preclinical safety testing of mAb is designed to identify possible toxicities in humans, to estimate the likelihood and severity of potential adverse events in humans, and to identify a safe starting dose and dose escalation, when possible. Preclinical testing concerns surrounding mAb products include their immunogenicity, stability, tissue cross-reactivity, and effector function(s). Species differences may complicate the design and interpretation of preclinical studies. CBER recognizes that animal models expressing the antigen of interest or a closely related, highly cross-reactive epitope are not always available. Pharmacokinetic and pharmacodynamic properties of mAb that are dependent upon specific antigen binding may not be evident in animal studies conducted in species which do not express the antigen of interest. In some cases, xenograft models can be developed by introducing cells expressing the antigen of interest into immunodeficient mice (e.g. SCID or nude mice). Such models can provide information on specific targeting of desired cells, especially with radiolabeled mAb or immunoconjugates. Transgenic models expressing the antigen of interest are another possibility, if available. Whenever they are available, parallel models which explore the effects of mAb against the animal homolog of the antigen of interest can be informative. *In vivo* activity models have proven valuable in providing data which support a rationale for the proposed product use and in defining safety and toxicity. Animal disease models are available to study the effects of mAb on many inflammatory and autoimmune diseases, and allograft rejection. The extent of preclinical safety testing and the results of such testing will influence safety considerations for initial clinical trials (e.g. starting dose, dose escalation scheme, etc.).

b. Preclinical testing schemes should parallel to the greatest extent feasible those anticipated for clinical use with respect to dose, concentration, schedule, route, and duration. The range of doses selected for study should include at least one dose that is equivalent to and one dose that is a multiple of the highest anticipated clinical dose, with appropriate adjustments for interspecies differences in body size. A broad dose range should be explored. The highest doses tested should elicit adverse effects, whenever possible. Dose ranges are best established with a minimum of three doses. The linearity and overall shape of the dose response curve should also be defined by investigation of several doses and dosing intervals. If changes in manufacturing and/or formulation are made subsequent to conduct of preclinical studies, the decision to repeat some or all preclinical studies should depend on an assessment of the impact or likely impact of these changes on the product (see Section II.E.).

2. Animal toxicology studies

When planning toxicity testing for mAb, the following should be considered:

a. If the test article is an unconjugated antibody and there is no animal model of disease activity or animal that carries the relevant antigen, and cross-reactivity studies with human tissues are clearly negative, toxicity testing may not be necessary.

b. When a relevant animal model is available, an attempt should be made to study the dose-dependence of pharmacodynamic effects. The use of a broad range of doses, including high doses may allow a better prediction of the therapeutic index.

c. The properties of a relevant antigen in the animal should be comparable to those in humans in

biodistribution, function, and structure. For example, studies of $CD34^+$ progenitor cells in the baboon are useful because the same cell fractions in both species express the $CD34^+$ antigen and produce hematopoietic engraftment. Absolute equivalence of antigen density or affinity for the mAb, however, is not necessary for an animal model to be useful. Differences in binding, for example, may be compensated for by alterations in the dose or dosing frequency. Differences between the animal and human in antigen number, the affinity of a mAb for the antigen, or the cellular response to mAb binding, should be identified. This will allow more accurate extrapolation of safe human starting dose and estimation of the margin of safety.

d. Routine assessments of mutagenicity are not generally needed for mAb.

e. Reproductive and developmental studies including teratogenicity in an appropriate animal species should be carried out in instances in which the product is intended for repeat or chronic administration to women of childbearing potential. Results of such peri- and post-natal developmental studies should be submitted for marketing approval. Evaluation of male fertility, when appropriate, should be completed before phase 3 trials.

3. Pharmacokinetics and pharmacodynamics

A pharmacokinetic model may aid in the interpretation of preclinical activity and toxicity, and in the recommendation of an appropriate dosing regimen and thereby improve the design of clinical trials. Such studies should aim at determining pharmacokinetic and pharmacodynamic endpoints. Of particular importance to the selection of clinical dosage is determining the relationship activity to area under the curve (AUC) of tissue or blood concentration over time. In considering the relationship of activity to AUC, factors related to the pharmacodynamics of the monoclonal antibody should be used in evaluating potential clinical effects. These factors include pathophysiologic status, threshold for effects as well as molecular events like the rates of association and dissociation for the site of action. Studies of biodistribution may provide the initial evidence for inappropriate tissue targeting by a mAb or explain toxicities that are observed in animals. Interpretation of data should consider species of origin, isotype, whether the mAb is an intact immunoglobulin, a fragment or an immunoconjugate, method of labeling, stability of the immunoconjugate, level of antigen expression in the recipient, binding to serum proteins, and route of administration. Even if antigen is expressed in an animal model, the mAb may bind the human target antigen and its animal counterpart with different affinities. MAb half-life may also be affected by glycosylation, susceptibility to proteases, presence of circulating antigen, and host immune response. The presence of antibodies to the product may alter biodistribution and elimination. In some cases, informative pharmacokinetic studies may be obtained in animal models which do not express specific antigenic determinants, depending upon the role played by antigen binding in product biodistribution, biotransformation and excretion.

a. Selection of the animal species for pharmacokinetic and pharmacodynamic testing should be guided by the following considerations:

i. Preference should always be given to study of a mAb in an animal model in a species that shares a cross-reactive or identical target antigen with humans, whenever such a species is available. For unconjugated mAb directed at human antigens not expressed in animal models or foreign antigens (bacterial, viral, etc.), studies in animal species lacking the target antigen may not be necessary unless they are designed to address manufacturing issues (see Section II.E. on Product Comparability).

ii. Study of non-human primates is appropriate for unconjugated mAb when there are antigen

binding data that indicate that primates are the most relevant species.

iii. Normal rodent and murine xenograft models should be critically evaluated for their likelihood of predicting accurately human pharmacokinetic behavior of mAb. Xenograft models may be more useful in evaluating the ability of mAb to bind to human tumors *in vivo*.

b. Changes in manufacturing or formulation may result in significant changes in biological activity. Therefore, it is recommended that the material used in the preclinical studies be manufactured using the same procedures as used or intended for use in manufacturing material for clinical trials. In some cases it may be appropriate to modify the components of the formulation for preclinical testing. For example, substitution of the homologous animal serum albumin for human serum albumin that is used as a carrier will prevent the formation of anti-albumin antibodies in animal studies and thereby increase the relevance of preclinical testing.

c. Pharmacokinetic parameters should be defined using one or more assay methods (e.g., a radiolabeled mAb should be assayed by ELISA and by measurement of radioactivity). In the case of immunoconjugates of any type, intact conjugate should be distinguished from free mAb and free ligand (e.g. toxin, drug, or radionuclide). Pharmacokinetic parameters that are most important for product characterization, as well as most useful for determining product comparability include T_{max}, C_{max}, $T_{1/2}$ and AUC.

d. The development of anti-immunoglobulin antibodies greatly complicates study and interpretation of the effects of repeat dosing in animals. Murine antibodies are non-immunogenic in mice but are immunogenic in humans, making it difficult to extrapolate the results of repeat dose studies in mice to planned repeat dose administration in humans. The reciprocal problem will occur with fully human, chimeric or "humanized" mAb. Repeat dose studies in rodents in this case may be of little value.

4. Preclinical *in vivo* studies with immunoconjugates

a. Immunoconjugates should be tested for stability *in vivo*.

i. Individual components of an immunoconjugate should be measured during pharmacokinetic and tissue distribution studies in animals and compared to the distribution of unconjugated antibody.

ii. The target tissues for the various components and the potential toxicities that they may cause should be established.

b. Immunoconjugates containing radionuclides, toxins, or drugs should undergo animal toxicity testing even when the target antigen is not present in an animal species, because of possible conjugate degradation or activity in sites that are not the result of mAb targeting. Depending upon the nature of the components of the immunoconjugate and the stability of the conjugate itself, separate studies of the components may be warranted. The toxicity profile of each component should adequately describe the incidence and severity of possible adverse effects. Results should be correlated closely with studies of conjugate stability. Studies of the immunoconjugate should be performed in a species with the relevant target antigen or disease model, whenever available and generally in rodents if a target antigen-positive species is not available. Toxicity testing of free toxin or nuclide may be performed in a different species.

c. For immunoconjugates containing radionuclides:

i. Animal biodistribution data may be used for initial human dose estimation.

ii. Animal models that express the targeted antigen, whenever such models are available, are more likely to reveal the effects of antigen "sinks" or tissues with unexpected antigen expression on biodistribution and/or toxicity.

iii. Xenograft models may evaluate tissue targeting and antigen non-specific radioimmunoconjugate distribution problems, but are not helpful at identifying areas of normal tissue cross-reactivity.

iv. An adequate number of animals should be studied to achieve radiation dose estimates with an acceptable coefficient of variation (usually less than 20%).

v. There should be complete accounting of the metabolism of the total dose of administered radioactivity and an adequate number of time points to determine early and late elimination phases.

vi. Radioimmunoconjugates should be tested for stability *in vitro* by incubation in serum or plasma (see Section II.C.7.). Methods should be developed to estimate the percent radioactivity in each of the three species of concern: free isotope, conjugated mAb, and labeled, non-mAb substances.

IV. CLINICAL STUDIES

MAb administered to humans have usually been well-tolerated. Instances of serious or fatal adverse events have generally resulted from intended or unintended binding of mAb to specific antigens. These events emphasize the importance of screening tissues for mAb cross-reactivity, particularly when relevant-antigen animal models are not available. The results of preclinical tests may alert physicians to potential toxicities and may indicate that more conservative dosing schemes are justified during dose escalation.

A. CLINICAL CONSIDERATIONS FOR PHASE 1 AND 2 STUDIES

1. General

a. Different approaches to Phase 1 and 2 studies may be warranted depending on the nature of the mAb. Initial studies of therapeutic mAb in phase 1 are generally escalation studies of single-doses of the mAb. The goal should be to determine a presumed optimal biologic dose (OBD) that is usually defined by pharmacokinetics or pharmacodynamic measurements (e.g., degree of antigen binding or saturation or target blood levels, determined on the basis of preclinical studies) and, where appropriate, by the tolerability to the agent (e.g., the maximally tolerated dose [MTD]) (24-26). In the case of a therapeutic unconjugated mAb, studies that identify the MTD may not be necessary. Instead, determination of a presumed OBD may be a more appropriate goal. Immune activation, when relevant to the mechanisms of action or toxicity of the mAb, should be evaluated. In the case of radiolabeled therapeutic mAb or immunotoxins, undesired tissue targeting and release of conjugate due to degradation are major safety concerns. Patients receiving immunotoxins should be monitored for capillary leak syndrome and for hepatic, renal, and muscle toxicities.

Some antibody-specific side effects are more likely to occur with certain subclasses of immunoglobulins.

These antibodies (e.g., human IgG1 and IgG3, mouse IgG2a) are more likely to activate complement or activate antibody-dependent cell-mediated toxicity (ADCC) via their Fc regions, leading to lysis of bound cells. MAb may also cause desired or adverse effects by blocking or inducing functions of target cells (e.g., cytokine release syndrome following stimulation of T-cell receptors by anti-CD3).

b. In general, subjects in clinical trials of therapeutic and diagnostic products, including mAb, should be representative of the population targeted for eventual product use. Because of the potential immunogenicity of mAb, healthy volunteers may not be appropriate candidates for phase 1 trials. The nature of the mAb, the target antigen, and the proposed clinical application should be considered before deciding to enroll healthy volunteers in a trial. Situations in which healthy volunteers might be used in early trials include the following:

i. When the risks of studying a new agent initially in the index population are too high, such as when the index population expresses abnormally high levels of antigen (raising specific toxicity concerns) or when the index population may be particularly vulnerable to toxicity because of serious illness or significant organ dysfunction.

ii. When the index population is so ill that safety data are confounded and difficult to interpret.

When healthy volunteers are considered for inclusion in initial studies of a mAb, the informed consent should reflect the absence of direct medical benefit. For healthy volunteers, as for patient volunteers, the informed consent process should also illustrate the potential immediate and long-term risks of receiving xenogeneic proteins. These include possible toxicity, allergic reactions, and, in the case of murine and other non-human mAb, potential future inability to receive or benefit from a diagnostic or therapeutic mAb because of the development of an immune response against the foreign protein.

c. Sponsors and investigators should carefully consider whether single doses of the mAb, multiple doses of the mAb in a single course, or multiple courses of therapy will be most likely to optimize benefit over risk. Concomitant therapies or repeat administration of the mAb may alter its safety and efficacy profiles. Changes in antigen mass (e.g., due to binding and clearance, or to antigen modulation by the mAb) and immune responses to the mAb, for example, may prevent extrapolation of single dose data to multiple-dose schedules. Furthermore, repeat administrations in the face of an antibody response against the therapeutic agent may lead to toxicity and/or loss of therapeutic benefit.

d. Subjects with prior parenteral exposure to xenogeneic proteins or with a history of xenogeneic protein allergies should be excluded from phase 1 studies of mAb products that have been derived from the same or a closely related species.

e. Informed consent issues in feasibility clinical trials in serious or immediately life-threatening conditions:
If applicable, informed consent forms for these trials should clearly state the following in language understandable to the patients:

i. only a limited characterization of the processes used to prepare the product for their ability to remove endogenous or exogenous infectious/toxic agent(s) was performed, and

ii. there may be potential health risks, including hitherto unknown risks, derived from exposure to such agents if they are present in the product.

2. Dose-setting

a. Whenever possible, the selection of the phase 1 starting dose should be based on safety and toxicity information derived from testing in a relevant animal model. When extrapolating from animal doses to human doses, information about the relative affinity of the mAb for the human antigen as compared to its animal analogue may be of great value. The target *in vivo* dose or concentration range should be based both on *in vitro* studies of cells for which antibody-antigen affinity and functional activity (e.g., immune modulation, cytotoxicity) have been measured, as well as on study of a relevant-antigen animal model, if available, to assess *in vivo* activity. If animal studies are judged to be impossible or of no relevance and initial *in vivo* studies are to be performed in humans, testing should begin at a low dose that is based on extrapolation from tissue culture studies or from available information gathered in clinical trials of a similar mAb.

b. Initial studies of radioimmunotherapeutic mAb should also employ escalating single-doses of the mAb, with the lowest and highest doses based on animal dosimetry and on the projected tolerance of normal organs to radiation. Both the elimination half-life of the mAb and the elimination half-life of the radioactivity should be characterized.

c. If a multiple-dose regimen of a mAb is anticipated, multiple-dose schedules should be explored in late phase 1 or phase 2 trials, after basic data on toxicity, peak levels, clearance, distribution, and biologic effects are available from single-dose studies. The time required for recovery from the biologic effects of single doses (e.g., immune recovery after $CD4^+$ cell depletion or modulation, return of bone marrow function after radioimmunotherapy) should also be well understood prior to initiation of multiple-dose regimens. The rationale for dosing schedules should be provided. The rationale should be based on dose tolerance, available pharmacokinetic and pharmacodynamic data in humans, and on relevant animal models of safety and efficacy. Modified dosing regimens to compensate for a high antibody response against the agent or circulating antigen may need to be studied. Pharmacokinetic studies to determine the relationships of human anti-mAb antibody titers and circulating antigen levels with organ distribution, clearance, and toxicity may be necessary.

d. Before repeat administration of a radioimmunotherapeutic or immunotoxin, the investigator should characterize all organ toxicities and pathology resulting from single dose administration. The timing of recovery from all toxic effects should be determined. Intra-patient dose escalation may confound interpretation of safety data because it may be difficult to determine whether toxic effects (e.g., to bone marrow) are due to prolonged therapy or to increased dosing levels. Intra-patient dose escalation may be appropriate if no toxicity is seen at the initial dose levels or if it is possible to use initial safe "test" doses and if cumulative toxicity is deemed unlikely. If intra-patient dose escalation is performed, consideration should be given to threshold and carryover effects, as well as to the reversibility of clinical and laboratory adverse events.

e. Design of pharmacokinetic studies should include consideration of the species in which the immunoglobulin is produced, the immunoglobulin class and subclass, and the structure of the antibody (e.g., whole mAb, Fab fragment) or immunoconjugate. A relevant study population will have the appropriate antigen and antigen mass. If antigen mass is likely to alter the bioavailability of the mAb, this should be determined in pharmacokinetic studies so that its impact on dose setting and on stratitifcation and analysis of later trials can be considered. Aside from obtaining estimates of common pharmacokinetic parameters, pharmacokinetic studies may also be very useful in situations in which the comparability of different products or formulations is to be determined (see Sections II.E. and III.A.). Pharmacokinetic studies optimally include the following:

i. Determination of plasma concentration profiles, distribution, and clearance of the mAb.

ii. Determination of doses for further study based on dose-concentration effect relationship and correlation with desired concentrations estimated from *in vitro* studies.

iii. Determination of peak and trough mAb levels and elimination rate constants.

iv. Determination of the organs and sites where the mAb is distributed, metabolized, and eliminated.

v. Determination of the fate of immunoconjugates by assaying the whole molecule and its components.

vi. Investigation of the relationships between the elimination rate and the method of administration, antigen load, and presence of a circulating antigen or of an antibody response against the therapeutic agent.

B. IMMUNOGENICITY: CLINICAL CONSIDERATIONS

Monitoring of antiglobulin titers and immune activity is of great importance in evaluating the safety and efficacy of mAb and in designing protocols involving repeat administration. Immune responses to mAb may have little or no effect, or may interfere significantly with the safety and/or efficacy of a mAb.

1. Monitoring the development of antibodies to mAb

Depending on the source of the mAb, assays for anti-immunoglobulins will need to be developed to detect human anti-mouse antibodies (HAMA), human anti-rat antibodies (HARA), human anti-human antibodies (HAHA), human anti-chimeric antibodies (HACA), and anti-idiotype antibodies. As appropriate for the mAb, assays should be developed to detect human immunoglobulins directed against humanized or primatized antibodies, immunonuclides and immunotoxins, their individual components (e.g., ricin), and neoantigens formed by the linked antibody/toxin/nuclide.

a. The timing of sample collection for evaluating the presence of an anti-mAb antibody should take into account whether the mAb is intended to be given as single or as multiple doses. Titers of the anti-mAb antibody should always be established at baseline to account for pre-existing antibodies (including anti-globulin or anti-conjugate antibodies, when appropriate), and also before readministration of the mAb. Post-administration samples may be drawn early (e.g., two weeks after administration), but should also be drawn at later times (e.g., at six-eight weeks).

b. The assay(s) used to detect the anti-mAb antibody should be standardized to the greatest extent possible. Aliquots of a "reference" preparation of antibody, e.g. anti-mouse antibody for a HAMA assay, with defined specificity from a human or primate source should be aliquoted and frozen to facilitate future intra- and inter-study comparisons. The reference preparation can also be used to establish a standard curve for routine testing. The assay(s) should be validated by establishing sensitivity, specificity, precision, and accuracy. Inhibition or competition studies with both negative and positive controls should be used to demonstrate that the assay detects antibodies to the mAb product. Studies should assess the range of reactivity of normal individuals and should evaluate potential interference by serum components such as bilirubin and lipids.

c. The specificity of the immune responses to the mAb should be identified and characterized in a sample of patients. These studies should establish whether the responses are generated against a heavy-chain isotype determinant, a light-chain, constant (C) region, variable (V) region, idiotypic epitope(s), immunoconjugate, or neoantigen. These data will demonstrate whether it is possible to use an anti-mAb antibody test with broad specificity (e.g., detecting human antibodies reactive with antigens of both heavy and light chain constant regions of all foreign immunoglobulin classes), or whether a more restricted anti-mAb antibody test that is idiotype-specific is necessary. In certain instances the anti-mAb antibody assay should include the actual mAb product as the detection antigen.

d. The choice of the appropriate assay for anti-mAb antibody depends on the proposed use and labeling of the product. Development and validation of the assay should accompany the clinical development of the new mAb. The results of the anti-mAb antibody testing should be correlated with product efficacy and adverse events.

e. A license application submitted for a mAb to be administered in a repeated dosing regimen should include a clinically available, validated test that reliably measures human antibody responses to the mAb, if an anti-mAb response may affect the safety, efficacy or dosing of the product. If a commercially available HAMA (or other anti-mAb antibody) test kit, is available, it may be used provided it has been demonstrated to reliably detect antibody response against the new mAb product. In most cases, humanized mAb require an assay specific for the product itself. If no appropriate anti-mAb antibody test is available, a properly validated test system should be developed by the sponsor.

2. Clinical consequences of immunogenicity

a. When a patient is found to have developed an antibody response against the therapeutic or diagnostic mAb, adverse events should be anticipated and appropriate precautions taken. MAb are generally given in facilities where acute resuscitative care is immediately available. The use of non-hospital settings for mAb administration (e.g., clinic or home) should be justified by clinical safety data. Vital signs should be observed closely for at least one hour after completion of the mAb administration. The possibility of delayed adverse effects from immune responses to mAb should be considered and reflected in the trial design, including appropriate clinical and laboratory testing.

b. Anaphylaxis, anaphylactoid and other immune reactions

i. True IgE-mediated anaphylaxis to whole mouse immunoglobulins is infrequent. It is theoretically possible that anti-human allotype responses of an allergic nature could occur but they have not been reported to be of clinical importance to date. If the mAb is conjugated to chelating agents or toxins, the likelihood of allergic reactions may be greater. In all cases, repeated administration of a mAb increases the likelihood of a hypersensitivity response. Immediate hypersensitivity reactions may range from mild to severe. Skin testing is not advised because it is a poor predictor of sensitivity to mouse immunoglobulins and can cause sensitization.

ii. Infusional reactions such as chills, rigors, aches, and low grade fever, are common during or immediately following mAb administration (the incidence is approximately 5% with several antibodies). The mechanism of these reactions is not clear. The frequency and intensity of such reactions can often be controlled by using slower infusion rates or by pre-medication.

iii. Serum sickness is unusual following mAb administration but has been described. Unlike anaphylaxis and infusional reactions, which occur during or immediately after antibody treatment,

serum sickness is delayed by several hours. The correlation between circulating levels of soluble antigen and immunocomplex-mediated adverse events such as serum sickness should be explored if such adverse events are observed.

c. Anti-mAb antibodies can interfere directly with some antibody-based clinical tests for antigens, such as CA-125 and CEA, by binding to the murine detection antibody. Indirect interference with diagnostic assays is theoretically possible if mAb administration induces anti-idiotypic responses that mimic the antigen. When appropriate, evidence for either type of interference should be systematically sought using well designed *in vitro* studies. Ideally, attempts should be made to circumvent such interference and alternative clinical assays should be validated.

d. When subjects are selected for testing a mAb, the risk that future therapy with a monoclonal antibody may be compromised by elicitation of an antibody response against the therapeutic agent or other mAb should be considered and reflected in the informed consent form.

C. PRODUCT-RELATED CONSIDERATIONS FOR PHASE 3 STUDIES

When planning manufacturing changes or scale-up programs during phase 3 clinical trials, sponsors should consider that product comparability may have to be demonstrated (see ref. 23 and Section II.E.). This may or may not require additional clinical studies depending upon the adequacy of preclinical data (see ref. 23 and Section II.E.). Thus, it is suggested that scale-up programs and contemplated changes in product manufacture be anticipated prior to the initiation of phase 3 trials. Sponsors should study a number of separate product lots during drug development to demonstrate that a safe and effective product can be prepared reliably.

D. ADMINISTRATION OF RADIOLABELED ANTIBODIES

1. Dosimetry

Grade 3 and grade 4 organ toxicities have been reported with therapeutic radioisotopes. Therefore, dosimetry estimates for human subjects are required prior to the initiation of phase 1 studies. The dosimetry estimates should be developed with simulation models utilizing an appropriate diagnostic radioisotope label on the antibody. If no diagnostic radiolabel for the antibody is available for simulation, animal studies with the therapeutic radiolabel may be utilized for dosimetry estimates. The actual dosimetry data for the therapeutic radiolabeled antibody, itself, should be acquired concurrent with the initial phase 1 study and reported prior to the initiation of a phase 2 study.

For diagnostic radiolabeled antibodies, as for the therapeutic isotopes, the investigator should provide estimates of the organ dosimetry prior to the first phase 1 study. Final dosimetry calculations from human studies should be completed prior to the submission of the license application.

a. General considerations: Sufficient data from animal or human studies should be submitted to the IND, to allow a reasonable calculation of radiation-absorbed dose to the whole body and to critical organs upon administration to a human subject [21 CFR 312.23(a)(10)(ii)]. See Appendix III for a list of organs to be included in dosimetry estimates.

The amount of radiation delivered by internal administration of radiolabeled antibodies should be calculated by internal radiation dosimetry. The absorbed fraction method of radiation dosimetry is described in two systems [21 CFR 361.1 (b)(3)(iv)]:

i. The Medical Internal Radiation Dose (MIRD) Committee of the Society of Nuclear Medicine

ii. The International Commission on Radiological Protection (ICRP).

The investigator should specify which methodology is used. The mathematical equations used to derive the radiation doses and the absorbed dose estimates should be provided. Sample calculations and all pertinent assumptions should be listed and submitted.

Safety hazards for patients and health care workers during and after administration of the radiolabeled antibody should be identified, evaluated, and managed appropriately.

b. Calculation of radiation dose to the target organ

Investigators should determine the following, based on the average patient:

i. The amount of radioactivity that accumulates in the target tissue/organ.

ii. The amount of radioactivity that accumulates in tissues adjacent to the target tissue/organ.

iii. The residence time of the radioactive mAb in the target tissue/organ and in adjacent regions.

iv. The radiation dose from the radioisotope, including the free radioisotope and any daughter products generated by decay of the radioisotope.

v. The total radiation from bound, free, and daughter radioisotopes associated with the radioimmunoconjugate, based upon immediate administration following preparation and upon delayed administration at the end of the allowed shelf life.

c. Maximum absorbed radiation dose

The amount of radioactive material administered to human subjects should be the smallest radiation dose that is practical to perform the procedure without jeopardizing the benefits obtained.

i. The amount of radiation delivered by the internal administration of radiolabeled antibodies should be calculated by internal radiation dosimetry using both the MIRD and ICRP systems. The higher estimate of the absorbed dose determined from either of these systems should be used in the radiation dosimetry safety assessment.

ii. Because of known or expected toxicities associated with radiation exposure, dosimetry estimates should be obtained as delineated in IV.D.1.a and b.

iii. Calculations should be provided that anticipate changes in dosimetry that might occur in the presence of diseases in organs that are critical in radioimmunoconjugate metabolism or excretion (e.g., renal dysfunction causing a larger fraction of the administered dose to be cleared via the hepatobiliary system or vice versa).

iv. Possible changes in dosimetry resulting from patient to patient variations in antigen mass should also be considered in dosimetry calculations (e.g., a large tumor mass may result in a larger than

expected radiation dose to a target organ from a radiolabeled anti-tumor mAb).

v. The mathematical equations used to derive the estimates of the radiation dose and the absorbed dose should be provided. Sample calculations and all pertinent assumptions should be listed.

vi. Calculations of dose estimates should be done assuming freshly labeled material to account for maximum amount of label as well as at the maximum shelf life of the radiolabeled antibody to allow for the upper limit of radioactive decay contaminants and should: (a) Include the highest amount of radioactivity to be administered; (b) Include the radiation exposure contributed by other diagnostic procedures such as roentgenograms or nuclear medicine scans that are part of the study; (c) Be expressed as Gray (Gy) per megaBequerel (MBq) or per millicurie of radionuclide; and (d) Be presented in a tabular format and include doses of individual absorbed radiation for the target tissues/organs and the organs listed in Appendix III.

2. Early clinical development of therapeutic radiolabeled mAb

a. Evaluations that should be conducted prior to Phase 1 studies.

Prior to phase 1 studies, investigators of therapeutic applications of radiolabeled antibodies should conduct the following evaluations for the average adult to be entered into the study:

i. The therapeutic radiolabeled antibody should be evaluated for *in vitro* stability and composition of the radioactive material to be administered. The expected and acceptable levels of the percent of free radioisotope, the percent of radioisotope bound to immunoreactive antibody, and the percent of bound radioisotope to non-immunoreactive antibody should be established. Calculations of the estimates should be at the maximum planned shelf life of the radiolabeled antibody to allow for the upper limit of radioactive decay contaminants and should be based upon the maximum dose of radioactive material to be administered to patients.

ii. The expected biodistribution and routes of clearance of the administered radiolabeled antibody dose fractions in tissues/organs should be defined.

iii. The expected biodistribution and routes of clearance that might occur in the presence of diseases in organs that are critical in radioimmunoconjugate metabolism or excretion should be described.

iv. The expected biodistribution and routes of clearance that might occur in the presence of immune responses (e.g., HAMA, HAHA, HARA) should be described.

With reference to the radioactive fractions of the administered radiolabeled antibody dose and the patterns of biodistribution, the following issues should be addressed:

v. From the biodistribution estimation, the expected residence time of the radiolabeled antibody fractions in the target tissues/organs and non-target tissues/organs should be determined.

vi. Based on the estimated residence times in each organ, the radiation exposure for each tissue/organ should be estimated.

vii. Based on the radiation exposure for each tissue/organ, the potential toxicity should be

described.

viii. Based on the potential radiation toxicity to tissues/organs, toxicity monitoring protocols should be developed and incorporated into the clinical trial.

ix. If the study has increasing doses of radioactive materials (e.g., a study of the maximum tolerated dose), the radiation exposure for tissues/organs and the associated potential toxicities should be estimated for each radiation dose level.

b. Selection of patients for phase 1 trials of therapeutic radiolabeled mAb.

Patients should be entered into phase 1 trials with therapeutic radiolabeled antibodies only after consideration of the following:

i. To reduce the potential for alterations in biodistribution, patients enrolled in early studies should not have prior or concurrent exposure to investigational or approved antibodies.

ii. Patients should be evaluated for immune response to the appropriate species of monoclonal antibody (e.g., HAMA, HARA). Patients demonstrating evidence of immune response should generally be excluded from phase 1 studies.

iii. For proper evaluation of potential adverse events in early studies, patients should be in adequate health to allow follow-up for three months without adjunctive chemotherapy or radiation therapy. Generally, patients should have a Karnofsky score greater than 70.

iv. Consultation with CBER is strongly recommended when considering the inclusion of pediatric patients in early radiolabeled antibody trials.

c. Study design issues for phase 1 studies of therapeutic radiolabeled mAb

Phase 1 studies should be designed to address the following points:

i. The prepared therapeutic radiolabeled antibody should be evaluated for *in vivo* stability. The previously estimated expected and acceptable levels of the percent of free isotope, the percent of bound radioisotope to immunoreactive antibody, and the percent of bound radioisotope to non-immunoreactive antibody should be confirmed and demonstrated to be reproducible. Calculations should be with the routine shelf life of the prepared radiolabeled antibody and based upon the maximum dose of radioactive material administered to patients.

ii. The pharmacokinetics and biodistribution in the patient population should be studied.

iii. The residence times with radiation dosimetry for tumor (when applicable), tissues, and organs should be determined.

iv. The pattern of toxicity, its relationship to administered dose and the organs of concern for acute and delayed radiation injury should be established.

v. Any apparent evidence of response of tumor to the administration of the radiolabeled antibody should be documented.

vi. The trial design should incorporate patient imaging with a diagnostic radiolabel on the antibody to confirm the expected biodistribution and residence times prior to the administration of the therapeutically radiolabeled antibody.

3. Adverse events for patients enrolled in trials of therapeutic radiolabeled mAb

The mechanism for follow-up of patients and reporting of adverse events should be described in the protocol prior to initiation of the trial. Complete evaluation and reporting of the adverse events potentially associated with the therapeutically radiolabeled antibody should be assured. If patients are referred to their attending physicians during the follow up phase, the investigator should plan and control the follow up of the treated patients for complete and timely reporting of adverse events potentially associated with the administration of the therapeutic radiolabeled antibody.

Patients removed from a trial should be continued in follow up for three months. All adverse events during that time interval should be reported, even if they are not thought to be related to the administered radiolabeled antibody.

4. Clinical development of radiolabeled mAb used as imaging agents

a. Prior to the initiation of phase 3 studies

Investigators of diagnostic applications of radiolabeled antibodies should collect stability, safety and pharmacokinetic information for the average adult expected to be entered into phase 3 studies.

i. The expected percent of free radioisotope, the expected percent of radioisotope bound to immunoreactive antibody, and the percent of the radioisotope bound to non-immunoreactive antibody should be determined. If unlabeled antibody is to be administered in conjunction with radiolabeled antibody, the ratio and amounts of the labeled and unlabeled antibody should be defined.

ii. The biodistribution of the administered radiolabeled antibody dose fractions in tissues/organs should be delineated, and from the biodistribution, the expected residence time of the radiolabeled antibody fractions in the target tissues/organs and non-target tissues/organs should have been estimated. The routes of clearance of the radiolabeled antibody should be determined.

iii. The changes in pharmacokinetics of the radiolabeled antibody with organ impairment, antigen load in the circulation, and tumor burden should be evaluated. The potential for clearance artifact to degrade patient imaging should be explored.

iv. Estimates of appropriate imaging times and techniques should be developed. Adjunctive imaging aids (e.g., enemas, emptying of the urinary bladder) should be evaluated.

v. Evidence of image quality should be gathered. The ability of the radiolabeled antibody to image known and/or occult disease should be documented. These data should be compared to imaging data obtained using standard diagnostic techniques whenever possible.The incidence of false positive localization of the radiolabeled antibody and the incidence of misinterpretation of the images to produce false positive and false negative interpretations should be explored. Disease specific factors (e.g., stage of disease, tumor burden, and co-morbid illness) should be evaluated for impact on technical procedures in the imaging protocol (e.g., time of imaging).

vi. Phase 2 trials should be designed to define the appropriate patient populations for phase 3 trial(s), to define the technical procedures used for imaging in the anticipated patient populations and to identify potential clinical utility of the test to be further explored in later studies.

vii. Multiple clinical sites should be employed in phase 2 studies to assess the reproducibility of the imaging techniques, and of the preparation and administration of the radiolabeled antibody.

b. Pivotal efficacy studies of radiolabeled mAb used as imaging agents

CBER staff should be consulted for review of and comments on Phase 3 protocol(s), prior to the initiation of the phase 3 study(ies). The following elements should be incorporated into each clinical protocol:

i. A prospectively defined and detailed primary efficacy endpoint and analytical plan.

ii. A study population consisting of those patients for whom the imaging agent is intended after licensure. The performance and utility of an imaging test may vary substantially based on the stage, extent, or severity of the disease, determined in part by the results of other diagnostic tests. Therefore, the study population and subpopulations to be analyzed should be carefully defined in terms of stage of disease as well as in terms of diagnostic tests performed and test results prior to study imaging. The protocols should be designed to assess the imaging performance and the utility in the populations.

iii. A plan for acquisition and storage of imaging data for radiolabeled antibody in the confirmatory studies.

iv. A prospective plan for evaluation of imaging performance:

(a). On-site image interpretation and reporting should be defined and documented. To the extent possible, the information available to the on-site reader should be defined by protocol and recorded on the case report forms.

(b). The off-site image interpretation should be the basis of the principal analysis of imaging performance in the phase 3 clinical trial. The off-site image interpretation and reporting of all radiolabeled antibody image findings and all confirmatory imaging should be defined and documented prospectively. The information available to the off-site reader should be defined by protocol and recorded on the case report form. In general, the off-site reader should have little information beyond the entry criteria of the study and specifically should not be aware of the on-site reading, the results of other diagnostic tests, or patient outcome data.

(c). Planned sample size should be sufficient to determine imaging performance measures to a predetermined precision (i.e., 95% confidence interval width). Imaging performance may vary with the stage, extent, and/or site of disease as defined by pre-imaging evaluations, and this should be accounted for prospectively in planning analysis of imaging performance. To determine performance, imaging results should be compared with another indicator of disease, usually results of standard imaging, biopsy, exploration, patient follow-up or some combination of these.

(d). If the planned use of the test is in conjunction with other diagnostic tests, its imaging performance should be determined and reported in groups of patients defined by the results of the other tests. For example, in some cases it will be important to know the imaging performance in patients with positive CT scans and the imaging performance in patients with negative CT scans.

v. A prospective plan for evaluation of clinical utility:

(a).When an agent has a significant incidence of false negative and false positives or significant toxicities, it is particularly important that the clinical utility be assessed to determine whether the value of the diagnostic information outweighs the potential adverse consequences of incorrect information or the toxicities. In this context, clinical utility means the extent to which information obtained by use of the mAb agent in a defined clinical setting can be expected to contribute to outcome, to contribute to the convenience or appropriateness of patient management, or to provide accurate prognostic information.

(b). Based on phase 2 and other data, the protocol should indicate the specific manner in which clinical utility is to be explored. The following issues should be addressed:

(1) The stage and severity of the disease in which the test is to be indicated should be specified.

(2) The protocol should specify whether the test is to be used in conjunction with or in lieu of other diagnostic tests. Because radiolabeled mAb image by a mechanism distinct from that of radiopharmaceuticals or diagnostic devices, the information obtained from a monoclonal antibody image may be complementary to that obtained by those means. For example, a mAb agent may not be as sensitive overall as an accepted standard test, but may be able to detect disease accurately under conditions where the standard technique fails.

(3) How the various results of the test are hypothesized to be clinically useful (for management or prognosis) should be clearly delineated. For example: positive results together with a positive CT scan are sufficiently diagnostic to avoid further diagnostic evaluation including biopsy; positive results are useful to guide biopsy; positive uptake is predictive of response to a specific therapeutic modality.

V. APPENDIX I: NORMAL HUMAN TISSUES USED IN CROSS-REACTIVITY TESTING

1. Adrenal
2. Bladder
3. Blood cells
4. Bone Marrow
5. Breast
6. Cerebellum
7. Cerebral cortex
8. Colon
9. Endothelium
10. Eye
11. Fallopian tube
12. Gastrointestinal tract
13. Heart
14. Kidney (glomerulus, tubule)
15. Liver
16. Lung
17. Lymph node
18. Ovary
19. Pancreas
20. Parathyroid
21. Pituitary
22. Placenta
23. Prostate
24. Skin
25. Spinal cord
26. Spleen
27. Striated muscle
28. Testis
29. Thymus
30. Thyroid
31. Ureter
32. Uterus (cervix, endometrium)

VI. APPENDIX II: MOUSE ANTIBODY PRODUCTION TEST

The following tests for murine viruses (mouse antibody production test (27) should be performed on any MCB and EPC derived from murine cell lines and on all lots of mAb derived from mouse ascites fluid:

1. Ectromelia
2. EDIM
3. GD VII virus
4. Hantaan virus
5. LCM virus, including challenge for non-lethal strains
6. LDH-elevating virus
7. Minute virus of mice
8. Mouse adenovirus
9. Mouse encephalomyelitis
10. Mouse hepatitis
11. Mouse salivary gland (murine CMV)
12. Pneumonia virus of mice
13. Polyoma
14. Reovirus type 3
15. Sendai
16. Thymic virus

VII. APPENDIX III: ORGANS TO BE CONSIDERED IN DOSIMETRY ESTIMATES

1. all target organs/tissues
2. bone
3. bone marrow
4. liver
5. spleen
6. adrenal
7. kidney
8. lung
9. heart
10. urinary bladder
11. gall bladder
12. thyroid
13. brain
14. gonads
15. gastrointestinal tract
16. adjacent organs of interest

VIII REFERENCES

1. Points to Consider in the Characterization of Cell Lines Used to Produce Biologicals, CBER, FDA, 1993.

2. Points to Consider in the Production and Testing of New Drugs and Biologicals Produced by Recombinant DNA Technology, FDA, CBER, April 10, 1985 and Supplement to the Points to Consider in the Production and Testing of New Drugs and Biologicals Produced by Recombinant DNA Technology: Nucleic Acid Characterization and Genetic Stability, CBER, FDA, April 6, 1992.

3. Elimination of Establishment License Application for Specified Biotechnology Products and Specified Synthetic Biologics. Final rule. (May 14, 1996; 61 Federal Register 24227).

4. Guidance for Industry for the Submission of Chemistry, Manufacturing and Controls Information for a Therapeutic Recombinant DNA-derived Product or a Monoclonal Antibody Product for In Vivo Use, CBER/CDER, FDA, August, 1996.

5. Points to Consider in the Production and Testing of Therapeutic Products for Human Use Derived from Transgenic Animals, CBER, FDA, 1995.

6. International Conference on Harmonisation; Final Guideline on Quality of Biotechnological Products: Analysis of the Expression Construct in Cells Used for Production of r-DNA Derived Protein Products, (February 23, 1996; 61 Federal Register 7006).

7. International Conference on Harmonisation; Final Guideline on Stability Testing of Biotechnological/Biological Products, (July 10, 1996; 61 Federal Register 36466).

8. Marcus-Sekura, CJ. Validation of removal of human retroviruses. In Developments in Biological Standardization, volume 75, "International Symposium on Virological Aspects of the Safety of Biological Products", edited by the International Association of Biological Standardization, pp. 133-143, Karger, New York, New York, USA, 1991.

9. Rubinstein, LJ and Stein, KE. Murine immune response to the Neisseria meningitidis Group C capsular polysaccharide. J Immunology 1988; 141:4352-56.

10. Lindmo, T, Boven, E, Cuttitta, F, et al. Determination of the immunoreactive fraction of radiolabeled monoclonal antibodies by linear extrapolation to binding at infinite antigen excess. J Immunol Meth 1984; 72:77-89.

11. Nisonoff, A, Lamoyi, E. Implications of the presence of an internal image of the antigen in anti-idiotype antibodies: possible application to vaccine production. Clin Immunol Immunopathol 1981;21:397-406.

12. Stein, KE and Epstein, S. [letter] Mol Immunol 1991;28:193.

13. Burtonboy, G, Delferriere, N, Bodeus, M, and Mousset, B. Absence of Viral Particles in Rat Hybridomas. An Ultrastructural Study In Rat Hybridomas and Rat Monoclonal Antibodies (Ed. H. Bazin), pp. 465-473, CRC Press, Boca Raton, Florida, USA.

14. Dinowitz, M, Lie, YS, Low, MA, et al. Recent studies on retrovirus-like particles in Chinese hamster

ovary cells. In Developments in Biological Standardization, volume 76, "Continuous Cell lines, An International Workshop on Current Issues", edited by the International Association of Biological Standardization, pp. 201-207, Karger, New York, New York, USA, 1992.

15. Miller, DG and Miller, AD. Tunicamycin treatment of CHO cells abrogates multiple blocks to retrovirus infection, one of which is due to a secreted inhibitor. J Virol 1992;66:78-84.

16. Eglitis, MA, Kadan, MJ, Wonilowicz, E, Gould, L. Introduction of human genomic sequences renders CHO-K1 cells susceptible to infection by amphotropic retroviruses. J Virol 1993;67:1100-1104.

17. Lander, M R and Chattopadhyay, S K, A Mus dunni cell line that lacks sequences closely related to endogenous murine leukemia viruses and can be infected by ecotropic, amphotropic, xenotropic and mink cell focus-forming viruses. J Virol. 1984;52:695-698.

18. Report of a WHO Study group (1987). Acceptability of cell substrates for production of biologicals. World Health Organization Technical Report Series 747.

19. Guideline on Validation of the Limulus Amebocyte Lysate Test as an End-Product Endotoxin Test for Human and Animal Parenteral Drugs, Biological Products, and Medical Devices, CDER/CBER/CDRH/CVR, FDA, December, 1987.

20. Guidelines on Sterile Drug Products Produced by Aseptic Processing, Center for Drugs and Biologics, FDA, June, 1987.

21. Guidance for Industry for the Submission Documentation for Sterilization Process Validation in Applications for Human and Veterinary Drug Products, CDER/CVM, FDA, November, 1994.

22. Guideline for Submitting Documentation for the Stability of Human Drugs and Biologics, Center for Drugs and Biologics, FDA, February, 1987.

23. FDA Guidance Concerning Demonstration of Comparability of Human Biological Products, Including Therapeutic Biotechnology-derived Products, CBER, FDA, April, 1996.

24. Goldsmith, MA, Slavik, M, Carter, SK. Quantitative prediction of drug toxicity in humans from toxicology in small and large animals. Cancer Res 1975;35:1354-64.

25. Collins, JM, Zaharko, DS, Dedrick, RL et al. Potential roles for preclinical pharmacology in phase I trials. Cancer Treat Rep 1986; 70:73-80.

26. EORTC Pharmacokinetics and metabolism group: pharmacokinetically-guided dose escalation in phase I clinical trials. Eur J Clin Oncol 1987; 23:1083-87.

27. Nicklas, W, Kraft, V, and B Meyer. Contamination of transplantable tumors, cell lines, and monoclonal antibodies with rodent viruses. Lab Anim Sci 1993;43: 296-300.

Guidance for Industry, Investigators, and Reviewers

Exploratory IND Studies

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

January 2006
Pharmacology/Toxicology

Contains Nonbinding Recommendations

Guidance for Industry, Investigators, and Reviewers

Exploratory IND Studies

Office of Training and Communication
Division of Drug Information, HFD-240
Center for Drug Evaluation and Research
Food and Drug Administration
5600 Fishers Lane
Rockville, MD 20857
(Tel) 301-827-4573
http://www.fda.gov/cder/guidance/index.htm

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

January 2006
Pharmacology/Toxicology

TABLE OF CONTENTS

Contains Nonbinding Recommendations

Guidance for Industry and Reviewers[1]

Exploratory IND Studies

This guidance represents the Food and Drug Administration's (FDA's) current thinking on this topic. It does not create or confer any rights for or on any person and does not operate to bind FDA or the public. Alternative approaches can be used if the approach satisfies the requirements of the applicable statutes and regulations. Discussions of an alternative approaches can be scheduled by contacting the FDA staff responsible for implementing this guidance. If the appropriate FDA staff cannot be located, contact can be made using the telephone number listed on the title page of this guidance.

I. INTRODUCTION

This guidance is intended to clarify what preclinical and clinical approaches, as well as chemistry, manufacturing, and controls information, should be considered when planning exploratory studies in humans, including studies of closely related drugs or therapeutic biological products, under an investigational new drug (IND) application (21 CFR 312). Existing regulations allow a great deal of flexibility in the amount of data that needs to be submitted with an IND application, depending on the goals of the proposed investigation, the specific human testing proposed, and the expected risks. The Agency believes that sponsors have not taken full advantage of that flexibility and often provide more supporting information in INDs than is required by regulations. This guidance is intended to clarify what manufacturing controls, preclinical testing, and clinical approaches can be considered when planning limited, early exploratory IND studies in humans.

For the purposes of this guidance the phrase *exploratory IND study* is intended to describe a clinical trial that

- is conducted early in phase 1,
- involves very limited human exposure, and
- has no therapeutic or diagnostic intent (e.g., screening studies, microdose studies).

---

[1] This guidance was developed by the Office of New Drugs in the Center for Drug Evaluation and Research (CDER).

This guidance contains information collection provisions that are subject to review by the Office of Management and Budget under the Paperwork Reduction Act of 1995 (44 U.S.C. 3501-3520). The collection of information in this guidance has been approved under OMB Control No. 0910-0014.

Such exploratory IND studies are conducted prior to the traditional dose escalation, safety, and tolerance studies that ordinarily initiate a clinical drug development program. The duration of dosing in an exploratory IND study is expected to be limited (e.g., 7 days). This guidance applies to early phase 1 clinical studies of investigational new drug and biological products that assess feasibility for further development of the drug or biological product.[2]

FDA's guidance documents, including this guidance, do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances means that something is suggested or recommended, but not required.

II. BACKGROUND

In its March 2004 *Critical Path Report*,[3] the Agency explained that to reduce the time and resources expended on candidate products that are unlikely to succeed,[4] new tools are needed to distinguish earlier in the process those candidates that hold promise from those that do not. This guidance describes some early phase 1 exploratory approaches that are consistent with regulatory requirements while maintaining needed human subject protection, but that involve fewer resources than is customary, enabling sponsors to move ahead more efficiently with the development of promising candidates.

A. Traditional Phase 1 Approach

Typically, during pharmaceutical development, large numbers of molecules are generated with the goal of identifying the most promising candidates for further development. These molecules are generally structurally related, but can differ in important ways. Promising candidates are often selected using in vitro testing models that examine binding to receptors, effects on enzyme activities, toxic effects, or other in vitro pharmacologic parameters; these tests usually require only small amounts of the drug. Candidates that are not rejected during these early tests are prepared in greater quantities for in vivo animal testing for efficacy and safety. Commonly, a single candidate is selected for an IND application and introduction into human subjects, initially healthy volunteers in most cases.

Before the human studies can begin, an IND must be submitted to the Agency containing, among other things, information on any risks anticipated based on the results of pharmacologic and

---

[2] Specifically, this guidance is limited to drug and certain well-characterized therapeutic biological products (e.g., recombinant therapeutic proteins and monoclonal antibodies) regulated by CDER. The guidance does not apply to human cell or tissue products, blood and blood proteins, vaccines, or to products regulated as devices.

[3] *Innovation or Stagnation, Challenge and Opportunity on the critical Path to New Medical Products* (March 2004).

[4] "A new medical compound entering phase 1 testing, often representing the culmination of upwards of a decade of preclinical screening and evaluation, is estimated to have only an 8 percent chance of reaching the market," *Critical Path Report*, March 2004.

toxicological data collected during studies of the drug in animals (21 CFR 312.23(a)(8)). These basic safety tests are most often performed in rats and dogs. The studies are designed to permit the selection of a safe starting dose for humans, to gain an understanding of which organs may be the targets of toxicity, to estimate the margin of safety between a clinical and a toxic dose, and to predict pharmacokinetic and pharmacodynamic parameters. These early tests are usually resource intensive, requiring significant investment in product synthesis, animal use, laboratory analyses, and time. Many resources are invested in, and thus wasted on, candidate products that subsequently are found to have unacceptable profiles when evaluated in humans — less than 10 percent of INDs for new molecular entities (NME) progress beyond the investigational stage to submission of a marketing application (NDA).[3] In addition, animal testing does not always predict performance in humans, and potentially effective candidates may not be developed because of resource constraints.

Existing regulations allow a great deal of flexibility in terms of the amount of data that need to be submitted with any IND application, depending on the goals of the proposed investigation, the specific human testing proposed, and the expected risks. The Agency believes that sponsors have not taken full advantage of that flexibility. As a result, limited, early phase 1 studies, such as those described in this guidance, are often supported by a more extensive preclinical database than is required by the regulations.

This guidance describes preclinical and clinical approaches, and the chemistry, manufacturing, and controls information that should be considered when planning exploratory IND studies in humans, including studies of closely related drugs or therapeutic biological products, under a single IND application (21 CFR 312).

B. Exploratory IND Approach

Exploratory IND studies usually involve very limited human exposure and have no therapeutic or diagnostic intent. Such studies can serve a number of useful goals. For example, an exploratory IND study can help sponsors

- Determine whether a mechanism of action defined in experimental systems can also be observed in humans (e.g., a binding property or inhibition of an enzyme)
- Provide important information on pharmacokinetics (PK)
- Select the most promising lead product from a group of candidates[5] designed to interact with a particular therapeutic target in humans, based on PK or pharmacodynamic (PD) properties

---

[5] For the purposes of this guidance, the term *candidate*, or *candidate product*, is used to describe a drug or biologic that is being tested in early exploratory studies under an IND. This guidance ***does not*** distinguish between a *drug product* and a *drug substance* as some other Agency guidances do.

(Most guidances use the term *drug product* to refer to a finished dosage form (e.g., tablet, capsule, solution) that contains an active drug ingredient generally, but not necessarily, in association with inactive ingredients, or a finished dosage form that does not contain an active ingredient but is intended to be used as a placebo. *Drug substance* usually refers to any component that is intended to furnish pharmacological activity or other direct effect

- Explore a product's biodistribution characteristics using various imaging technologies

Whatever the goal of the study, exploratory IND studies can help identify, early in the process, promising candidates for continued development and eliminate those lacking promise. As a result, exploratory IND studies may help reduce the number of human subjects and resources, including the amount of candidate product, needed to identify promising drugs. The studies discussed in this guidance involve dosing a limited number of subjects with a limited range of doses for a limited period of time.

Existing regulations provide more flexibility with regard to the preclinical testing requirements for exploratory IND studies than for traditional IND studies. However, sponsors submitting the kinds of studies described in this guidance have not always taken full advantage of that flexibility. Sponsors often provide more supporting information in their INDs than is required by the regulations. Because exploratory IND studies involve administering either sub-pharmacologic doses of a product, or doses expected to produce a pharmacologic, but not a toxic, effect, the potential risk to human subjects is less than for a traditional phase 1 study that, for example, seeks to establish a maximally tolerated dose. ***Because exploratory IND studies present fewer potential risks than do traditional phase 1 studies that look for dose-limiting toxicities, such limited exploratory IND investigations in humans can be initiated with less, or different, preclinical support than is required for traditional IND studies.***[6]

The Agency expects that this early phase 1 exploratory IND approach will apply to a number of different study paradigms. Although his guidance explores several potential applications, many others can be proposed. The Agency believes that, consistent with its Critical Path Initiative, clarifying Agency thinking about how much and what kind of testing is needed to support early studies in humans will facilitate the entry of new products into clinical testing and speed product development.

Although exploratory IND studies may be used during development of products intended for any indication, it is particularly important for manufacturers to consider this approach when developing products to treat serious diseases. Because the approach can help identify promising candidates more quickly and precisely, exploratory IND studies could become an important part of the armamentarium when developing drug and biological products to treat a serious or life-threatening illness. The Agency has previously articulated its commitment to ensuring that appropriate flexibility is applied when patients with a serious disease and no satisfactory alternative therapies are enrolled in a trial with therapeutic intent.[7]

---

in the diagnosis, cure, mitigation, treatment, or prevention of disease, or to affect the structure or any function of the body.)

[6] Generally, these types of studies would not be carried out in pediatric patients or in pregnant or lactating women.

[7] Subpart H Accelerated Approval of New Drugs for Serious or Life-Threatening Illnesses. See also FDA guidance for industry *Fast Track Drug Development Programs — Designation, Development, and Application Review*.

III. CONTENT OF IND SUBMISSIONS

To begin any kind of testing in humans, applicants must submit an IND application to the Agency with certain types of information (see 21 CFR 312.23 IND Content and Format). The primary purpose of the IND submission is to ensure that subjects will not face undue risk of harm. The major information that must be submitted includes:

- Information on a clinical development plan
- Chemistry, manufacturing, and controls information
- Pharmacology and toxicology information
- Previous human experience with the investigational candidate or related compounds, if there is any

The following sections discuss the first three in more detail. Because the exploratory IND studies addressed by this guidance will be first in human studies, previous human experience is not pertinent and will not be discussed. The common theme throughout is that, depending on the study, the informational requirements for exploratory IND studies are more flexible than for traditional IND studies.

A. Clinical Information

1. Introductory statement and general investigational plan

A traditional IND application describes the rationale for the proposed clinical trial program and discusses the potential outcome of the clinical investigation. The exploratory IND studies discussed here focus on a circumscribed study or group of studies, and plans for further development cannot be formulated without the results of these studies. Therefore, an exploratory IND application should articulate the rationale for selecting a compound (or compounds) and for studying them in a single trial or related trials, as this represents all that is known about the overall development plan at this stage. This section should also make it clear that the IND is intended to be withdrawn[8] after completion of the outlined study or studies.

2. Types of studies

Potentially useful study designs include both single- and multiple-dose studies. In single-dose studies, a sub-pharmacologic[9] or pharmacologic dose is administered to a limited number of subjects (healthy volunteers or patients). For example, microdose studies usually involve the single administration of a small dose with the goal of collecting pharmacokinetic information or performing imaging studies, or both.

---

[8] The withdrawn, or inactive, IND can be referenced in any subsequent traditional IND.

[9] A radiolabeled candidate compound can be administered at doses that are known to have no pharmacologic effect in humans without an IND application in basic research studies when the compound has previously been studied in humans and the results published in the literature. These basic research investigations are conducted under the oversight of an institutional review board (IRB) and a radioactive drug research committee (21 CFR 361.1).

Repeat dose clinical studies can be designed with pharmacologic or pharmacodynamic endpoints. In exploratory IND studies, the duration of dosing should be limited (e.g., 7 days). For escalating dose studies done under an exploratory IND, dosing should be designed to investigate a pharmacodynamic endpoint, not to determine the limits of tolerability.

B. Chemistry, Manufacturing, and Controls Information

The regulations at 21 CFR 312.23(a)(7)(i) emphasize the graded nature of chemistry, manufacturing, and controls (CMC) information needed as development under an IND application progresses. Although in each phase of a clinical investigational program sufficient information should be submitted to ensure the proper identification, strength, quality, purity, and potency of the investigational candidate, the amount of information that will provide that assurance will vary with the phase of the investigation, the proposed duration of the investigation, the dosage form, and the amount of information already available. For the purpose of an exploratory IND application, the CMC information indicated below can be provided in a summary report to enable the Agency to make the necessary safety assessment.

The sponsor must state in the beginning of the exploratory IND application whether it believes the chemistry or manufacturing of the candidate product presents any potential for human risk (e.g., specific findings in preclinical studies associated with known risks of related compounds) (§ 312.23). If so, these potential risks should be discussed, and the steps proposed to monitor for such risks should be described.

The Agency is in the process of developing guidance explaining the stepwise approach to meeting current good manufacturing practice (CGMP) regulations. Once finalized, that guidance will be useful to persons seeking to manufacture, or prepare, products intended for use in an exploratory IND study.

1. General information for the candidate product

Except as noted below, the extent and type of chemistry and manufacturing information to be submitted in an exploratory IND application is similar to that described in current guidance for use of investigational products.[10] Information on each candidate product (i.e., the active ingredient) can be submitted in a summary report containing the following items.

- Description of the candidate product, including physical, chemical, and/or biological characteristics, as well as its source (e.g., synthetic, animal source, plant extract, or biotechnology-derived) and therapeutic class (e.g., radiopharmaceutic, immunosuppressant, agonist, antagonist) (see sections below for exceptions).
- Description of the dosage form and information related to the dosage form

[10] See guidance for industry *Content and Format of Investigational New Drug Applications for Phase 1 Studies of Drugs, Including Well-Characterized, Therapeutic, Biotechnology-Derived Products.*

- Description of the formulation or routes of administration intended to be used in the human trial. For oral administration, sponsors can consider using suspensions or solutions in addition to the more usual tablets, powders, and capsules. For products intended for ophthalmic, inhalational (aqueous base), or parenteral administration, sterility and apyrogenicity must be ensured. For biological candidate products, freedom from contaminants associated with their manufacture, such as viruses, mycoplasma, and foreign DNA, also should be ensured. All excipients should be generally recognized as safe[11] or part of a formulation that is approved or licensed in the United States for the same route of administration and amount,[12] or adequately qualified through appropriate animal studies.
- The grade and quality (e.g., USP, NF, ACS) of excipients used in the manufacture of the investigational candidate product, including both those components intended to appear in the product and those that may not appear, but that are used in the manufacturing process
- Name and address of the manufacturer(s) (if different from the sponsor)
- The method of preparation of the candidate product lots used in preclinical studies and intended for the proposed human study, including a brief description of the method of manufacture and the packaging procedure, as appropriate, with a description of the container and closure system. For the active substance, include a list of the starting materials, reagents, solvents, catalysts used, and purification steps employed to prepare the candidate product. For sterile products, describe the sterilization process and controls for ensuring sterility. For biological/biotechnology-derived products, also identify the source material (e.g., Master Cell Bank), describe the expression system (e.g., fermentation methods) and harvest methods, as well as methods for removal/inactivation of potential viral contaminants. We recommend the use of a detailed flow diagram that includes all materials used as the usual, most effective, presentation of this information.
- Quantitative composition of the product
- A brief description of adequate test methods used to ensure the identity, strength, quality, purity, and potency accompanied by the test results, or a certificate of analysis, of the candidate product lots used in toxicological studies and intended for the proposed human study. For biotechnology products produced in mammalian cells or animals, this will include tests and studies to ensure the removal and/or inactivation of potential viral contaminants.
- Information that demonstrates the stability of the product during toxicology studies and an explanation of how stability will be evaluated during the clinical studies

---

[11] Excipients considered to be generally recognized as safe (GRAS) are included in a list that is maintained on the Internet at http://www.accessdata.fda.gov/scripts/cder/iig/index.cfm. See also 21 CFR 330.1, which explains the GRAS concept.

[12] Novel excipients should be appropriately qualified for their intended use. FDA has issued guidance on *Nonclinical Studies for Development of Pharmaceutical Excipients*.

- For ophthalmic, inhalational (aqueous base), or parenteral dosage forms, results from sterility and pyrogenicity tests

2. Analytical characterization of candidate product

There are two scenarios under which CMC information can be provided to an IND application. In the first scenario, the ***same batch*** of candidate product is used in both the toxicology studies and clinical trials. This material will be qualified for human use based on the CMC information (see III.B.1, above) and results of the toxicology studies described elsewhere in this guidance. Although we recommend establishing the impurity profile to the extent possible for future reference and/or comparison, not all impurities of the candidate product may need characterization at this stage of product development. If an issue arises during the toxicology qualification of the product, the appropriate parameters can be studied further, on an as-needed basis. Impurities (e.g., chemical and microbiological) should be characterized in accordance with recommendations in Agency guidance,[13] if, and when, the sponsor files a traditional IND for further clinical investigation.

In the second scenario, the batch of candidate drug product to be used in the clinical studies may not be the same as that used in the nonclinical toxicology studies. In such a case, the sponsor should demonstrate by analytical testing that the batch to be used is ***representative*** of batches used in the nonclinical toxicology studies. To achieve this, relevant analytical quality test results should be sufficient to enable comparison of different batches of the product. Tests to accomplish this include:

- Identity
- Structure (e.g., optical rotation (for chiral compounds), reducing/non-reducing electrophoresis (for proteins))
- Assay for purity
- Impurity profile (e.g., product- and process-related impurities, residual solvents, heavy metals)
- Assay for potency (biologic)
- Physical characteristics (as appropriate)
- Microbiological characteristics (as appropriate)

C. Safety Program Designs — Examples

Pharmacology and toxicology information is derived from preclinical safety testing performed in animals and in vitro. Preclinical studies for small molecules are described in ICH M3 while those for biologics follow guidance described in ICH S6. Some of the toxicology tests described in this guidance may not be appropriate for biologics. The toxicology evaluation recommended for an exploratory IND application is more limited than for a traditional IND application.[14] The

[13] See footnote 10 and guidance for industry, *INDs for Phase 2 and Phase 3 Studies, Chemistry, Manufacturing, and Controls Information.*

[14] International Conference on Harmonisation (ICH) guidance for industry *M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals* describes what is expected for a traditional IND.

basis for the reduced preclinical package is the reduced scope of an exploratory IND clinical study. Although exploratory IND studies in some cases are expected to induce pharmacologic effects, they are not designed to establish maximally tolerated doses. Furthermore, the duration of drug exposure in exploratory IND studies is limited. The level of preclinical testing performed to ensure safety will depend on the scope and intended goals of the clinical trials.

There are a number of study objectives for which the preclinical safety programs may be tailored to the study design. Examples include: confirming that an expected mechanism of action can be observed in humans; measuring binding affinity or localization of drug; assessing PK and metabolism; comparing the effect on a potential therapeutic target with other therapies. Three examples are discussed in detail in the following paragraphs.

1. Clinical studies of pharmacokinetics or imaging

Microdose studies are designed to evaluate pharmacokinetics or imaging of specific targets and are designed not to induce pharmacologic effects. Because of this, the risk to human subjects is very limited, and information adequate to support the initiation of such limited human studies can be derived from limited nonclinical safety studies. A *microdose* is defined as less than $1/100^{th}$ of the dose of a test substance calculated (based on animal data) to yield a pharmacologic effect of the test substance with a maximum dose of ≤100 micrograms (for imaging agents, the latter criterion applies).[15] Due to differences in molecular weights as compared to synthetic drugs, the maximum dose for protein products is ≤30 nanomoles.

FDA currently accepts the use of extended single-dose toxicity studies in animals to support single-dose studies in humans. For microdose studies, a single mammalian species (both sexes) can be used if justified by in vitro metabolism data and by comparative data on in vitro pharmacodynamic effects. The route of exposure in animals should be by the intended clinical route. In these studies, animals should be observed for 14 days post-dosing with an interim necropsy, typically on day 2, and endpoints evaluated should include body weights, clinical signs, clinical chemistries, hematology, and histopathology (high dose and control only if no pathology is seen at the high dose). The study should be designed to establish a dose inducing a minimal toxic effect, or alternatively, establishing a margin of safety. To establish a margin of safety, the sponsor should demonstrate that a large multiple (e.g., 100X) of the proposed human dose does not induce adverse effects in the experimental animals. Scaling from animals to humans based on body surface area can be used to select the dose for use in the clinical trial. Scaling based on pharmacokinetic/pharmacodynamic modeling would also be appropriate if such data are available.

Because microdose studies involve only single exposures to microgram quantities of test materials and because such exposures are comparable to routine environmental exposures, routine genetic toxicology testing is not needed. For similar reasons, safety pharmacology studies are also not recommended.

---

[15] See European Medicines Agency (EMEA), Evaluation of Medicines for Human Use, "Position Paper on Non-Clinical Safety Studies to Support Clinical Trials with a Single Microdose," CPMP/SWP/2599/02Rev 1, 23 June 2004.

2. Clinical trials to study pharmacologically relevant doses

A second example involves clinical trials designed to study pharmacologic effects of candidate products. More extensive preclinical safety data would be needed to support the safety of such studies. However, since the goal would not include defining a maximally tolerated dose, the evaluation can still be less extensive than typically needed to support a traditional IND application. See the flow chart in the Attachment to this document.

Repeat dose clinical trials lasting up to 7 days can be supported by a 2-week repeat dose toxicology study in a sensitive species accompanied by toxicokinetic evaluations. The goal of such a study would be to select safe starting and maximum doses for the clinical trial. The rat is the usual species chosen for this purpose, but other species might be selected. In addition to studies in a rodent species, additional studies in nonrodents, most often dogs, can be used to confirm that the rodent is an appropriately sensitive species. If it is known that a particular species is most appropriate for a class of compounds, studies can be limited to that species. This confirmation can be approached in a number of ways. A lack of gender difference in the rodent study can serve as a basis for testing only a single sex in the second species if only a single sex will be studied in the clinical trial.

The numbers of animals used in the confirmatory study can be fewer than normally used to attain statistically meaningful comparisons, but of sufficient number to rule out any toxicologically significant difference in sensitivity compared with rodent (e.g. four non-rodents per treatment group). The confirmatory study could be a dedicated study involving repeat administrations of a single dose level approximating the rat NOAEL[16] calculated on the basis of body surface area. Alternatively, the test in the second species could be incorporated as part of an exploratory, dose escalating study culminating in repeated doses equivalent to the rat NOAEL. The number of repeat administrations at the rat NOAEL should, at a minimum, be equal to the number of administrations, given with the same schedule, intended clinically. The route of administration should be the same as the expected clinical route, and toxicokinetic measurements should be used to assess exposure. The same endpoints assessed in the rodent study should be evaluated in the second species. If the data from the confirmatory study suggest that the rodent is not the more sensitive species, a 2-week repeated dose toxicity study should be performed in the second species to select doses for human trials. This study should include measurements of body weight, clinical signs, clinical chemistries, hematology, and histopathology.

In contrast to microdose studies, for clinical trials designed to evaluate higher or repeated doses, each candidate product to be tested should be evaluated for safety pharmacology.[17] Evaluation of the central nervous and respiratory systems can be performed as part the rodent toxicology studies while safety pharmacology for the cardiovascular system can be assessed in the nonrodent species, generally the dog, and can be conducted as part of the confirmatory or dose-escalation study.

[16] No-observed-adverse-effect level (NOAEL).

[17] For details see the guidance for industry *S7A Safety Pharmacology Studies for Human Pharmaceuticals*.

In general, each product in this type of exploratory IND should be tested for potential genotoxicity unless such testing is not appropriate for the population (e.g. terminally ill patients) or product to be studied. The genetic toxicology tests should include a bacterial mutation assay using all five tester strains with and without metabolic activation[18] as well as a test for chromosomal damage either in vitro (cytogenetics assay or mouse lymphoma thymidine kinase gene mutation assay) or in vivo. The in vivo test can be a micronucleus assay performed in conjunction with the repeated dose toxicity study in the rodent species. The high dose in this case should be a maximally tolerated or limit dose.

The results from the preclinical program can be used to select starting and maximum doses for the clinical trials. The starting dose is anticipated to be no greater than 1/50 of the NOAEL from the 2-week toxicology study in the sensitive species on a mg/m^2 basis. The maximum clinical dose would be the lowest of the following:

- ¼ of the 2-week rodent NOAEL on a mg/m^2 basis
- Up to ½ of the AUC at the NOAEL in the 2-week rodent study, or the AUC in the dog at the rat NOAEL, whichever is lower
- The dose that produces a pharmacologic and/or pharmacodynamic response or at which target modulation is observed in the clinical trial
- Observation of an adverse clinical response

Escalation from the proposed maximal clinical dose should only be performed after consultation with and concurrence of the FDA.

It is recognized that the studies described above are most appropriate for chemical drugs. Other animal models (e.g. nonhuman primates) may be more appropriate for biologics, and some tests may be inappropriate (e.g. genetic toxicology testing) for proteins.

3. Clinical studies of MOAs related to efficacy

A third example involves clinical studies intended to evaluate mechanisms of action (MOAs). To support this approach, the FDA will accept alternative, or modified, pharmacologic and toxicological studies to select clinical starting doses and dose escalation schemes. For example, short-term, modified toxicity or safety studies in two animal species based on a dosing strategy to achieve a clinical pharmacodynamic endpoint can in some instances serve as the basis for selecting the safe clinical starting dose for a new candidate drug. These animal studies would incorporate endpoints that are mechanistically based on the pharmacology of the new chemical entity and thought to be important to clinical effectiveness. For example, if the degree of saturation of a receptor or the inhibition of an enzyme were considered possibly related to effectiveness, this parameter would be characterized and determined in the animal study and then used as an endpoint in a subsequent clinical investigation. The dose and dosing regimen determined in the animal study would be extrapolated for use in the clinical investigation. In some cases, a single species could be used if it were established as the most relevant species based on scientific evidence using the specific candidate intended for the clinical investigation.

[18] *For* details see guidance for industry *S2A: Guidance on Specific Aspects of Regulatory Genotoxicity Tests for Pharmaceuticals* and *S2B: Genotoxicity: A Standard Battery for Genotoxicity Testing for Pharmaceuticals*.

Although the production of frank toxicity is not the primary intended goal of the nonclinical study, relevant informative endpoints (e.g., hematology and histopathology) selected as important for clinical safety evaluation should be investigated. For example, an antibody that binds with a high degree of selectivity to a tumor-associated antigen could be studied in accordance with this third category. The mechanism of action of antibody-based products is generally associated with their binding properties and the effect on functions associated with immunoglobulins. Pharmacology and toxicology studies provide information about the selection of doses used in clinical studies through evidence of both a safe upper and potentially efficacious lower limit of exposure. These doses might be consistent with target plasma levels of the drug based on animal models of disease. The upper safe levels could be established in animal studies that show a lack of toxicity at these levels.

D. GLP Compliance

It is expected that all preclinical safety studies supporting the safety of an exploratory IND application will be performed in a manner consistent with good laboratory practices (GLP) (21 CFR Part 58). The GLP provisions apply to a broad variety of studies, test articles, and test systems. Sponsors are encouraged to discuss any need for an exemption from GLP provisions with the FDA prior to conducting safety related studies, for example, during a pre-IND meeting. Sponsors must justify any nonconformance with GLP provisions (21 CFR 312.23(a)(8)(iii)).

IV. CONCLUSION

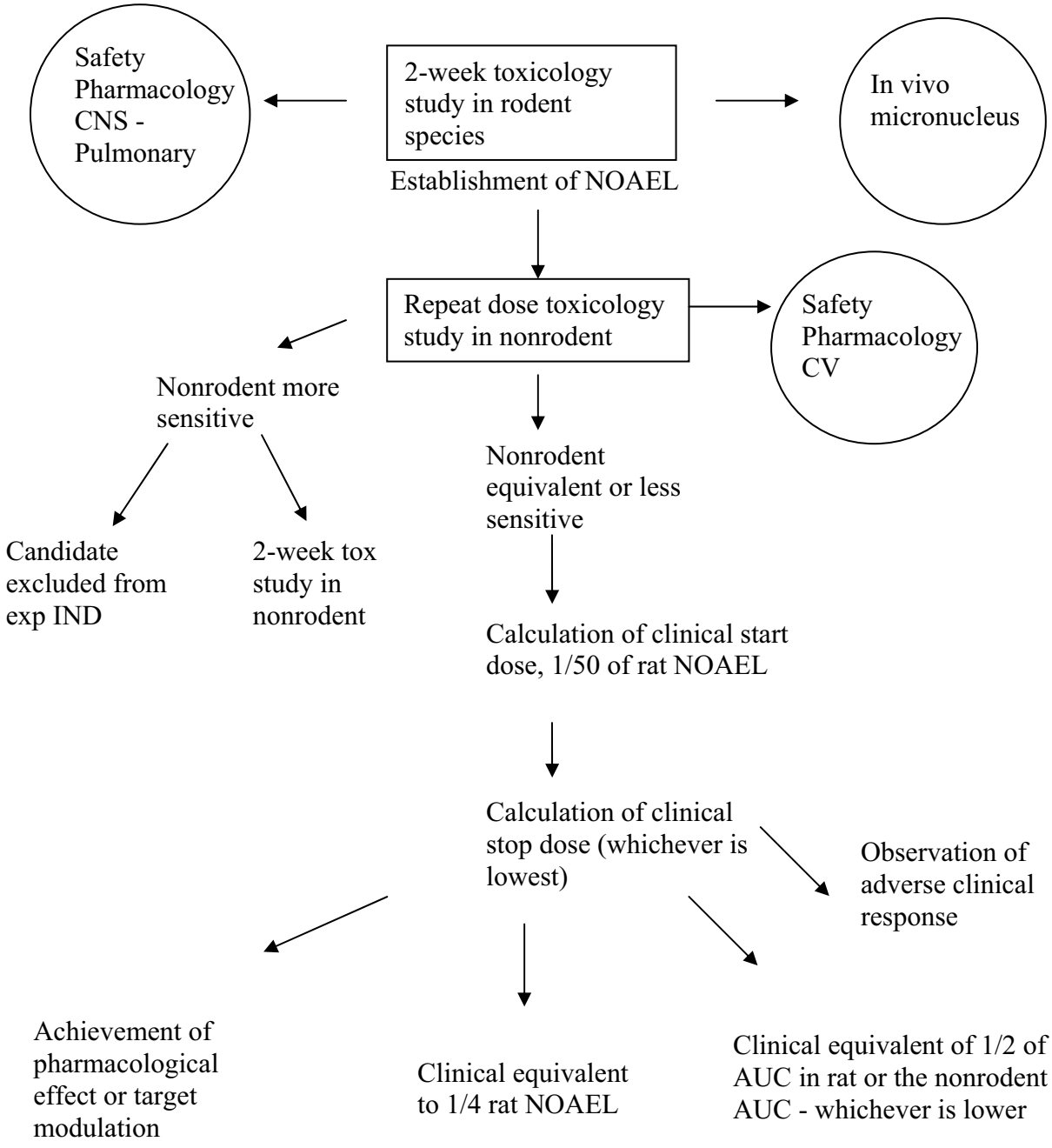
Existing regulations allow a great deal of flexibility in the amount of data that needs to be submitted with any IND application, depending on the goals of an investigation, the specific human testing being proposed, and the expected risks. Sponsors have not taken full advantage of that flexibility, and limited, early phase 1 studies, such as those described in this guidance, are often supported by a more extensive preclinical database than is needed for those studies alone.

The common theme throughout this guidance is that, depending on the study, the preclinical testing programs for exploratory IND studies can be less extensive than for traditional IND studies. This is because for the approaches discussed in this guidance, which involve administering sub-pharmacologic doses of a candidate product or products, the potential risks to human subjects are less than for a traditional phase 1 study.

The Agency is undertaking a number of efforts to reduce the time spent in early drug development on products that are unlikely to succeed. This guidance describes some exploratory approaches that are consistent with regulatory requirements, but that will enable sponsors to move ahead more efficiently with the development of promising candidate products while maintaining needed human subject protections.

ATTACHMENT

A Preclinical Toxicology Testing Strategy for Exploratory INDs Designed To Administer Pharmacologically Active Doses



Introduction

To facilitate the discussion of biological oncology products, the general approach to safety assessment based on nonclinical studies is provided below. Although a number of aspects of nonclinical testing are of concern to the FDA, sponsors and patients, of particular importance is the duration of nonclinical testing relative to that performed in support of early clinical investigations. Fundamentally, the duration of nonclinical studies for biologic oncology products varies in length in proportion to the planned duration of clinical study. However, it is recognized that in area of oncology, the duration of nonclinical studies may be significantly shorter in comparison to other diseases with less severe morbidity and mortality. Nevertheless in some few instances, nonclinical toxicology studies of 3 months rather than 1 month for novel therapeutics have been requested in which the underlying pharmacological and potential pathophysiological effects appear to be warranted based on approximating steady-state conditions, concerns regarding cumulative ablation of organ functionality or due to the potential emergence of new, significant toxicities.

As a future guidance in the area of nonclinical testing of biological oncology products is planned, we wish to fully consider a variety of perspectives on this issue. In addition to the text provided below that describes the fundamental process of nonclinical assessment for biological oncology products, an abstract of the presentation “Pre-Clinical Requirements for Phase 1 Studies – Biological Oncology Products” will be found. The FDA presentation on this subject will include the aforementioned presentation and one that describes the specific occurrence of new adverse findings observed in nonclinical studies of longer duration.

GENERAL PRINCIPLES OF NONCLINICAL TESTING FOR BIOLOGICAL ONCOLOGY PRODUCTS

Underlying the nonclinical assessment for the safety of biological products is a set of general principles which provide a flexible approach for sponsors to obtain data that should be considered in support of clinical studies in oncology. These principles highlight the selection of relevant animal model(s) and duration and frequency of dosing in nonclinical toxicology studies. Other aspects pertinent to quality and interpretability of the nonclinical studies include comparability of material used in nonclinical and clinical studies and compliance with Good Laboratory Practices (GLP). For nonclinical studies to have the greatest potential to guide future clinical investigations, the nonclinical study designs for toxicity testing should reasonably approximate the proposed clinical trials in terms of the anticipated dose ranges, schedules and routes of administration as well as duration of dosing. When these factors are appropriately considered and adjusted for interspecies differences, nonclinical toxicology studies provide a valid scientific means of selecting the initial dose, identifying target organs for toxicity, clinical monitoring strategies and the anticipation of potential risk.

Decisions regarding initial investigational new drug (IND) applications require a careful weighing of risk that is often based solely on nonclinical studies of pharmacology and toxicology and the potential for gathering useful clinical information in the proposed study. An evaluation of the desired pharmacologic effects, evidence from 'proof-of-principle' studies, and supportive nonclinical toxicology data are essential for making sound decisions about the clinical development. Primary outcomes of data derived from nonclinical pharmacology and toxicology studies are the clinical dosage(s), dosing schedules and monitoring. Informative nonclinical studies for biological products depend on the selection of a relevant animal model as significant differences may exist between species in their responsiveness to biological products. Once a relevant animal model is selected a variety of in vitro and in vivo studies may be performed that examine the pharmacology and toxicology of the biological product (Guidance for Industry S6 Preclinical Safety Evaluation of Biotechnology-Derived Pharmaceuticals, 1997).

Relevant Animal Model

Nonclinical studies should be conducted in relevant species to provide information useful to evaluating risk to patients or subjects in the clinical study. A relevant species is one in which the compound has pharmacological activities similar to that of humans due to the expression of a responsive receptor, in the case of cytokines, or an epitope, in the case of mAbs. In the case of mAbs, the definition of a relevant animal model may include animals that demonstrate a tissue cross-reactivity profile similar to humans. Ideally, for toxicity testing of monoclonal antibodies (mAbs), the animal model selected will express the desired epitope and demonstrate a tissue cross-reactivity profile similar to that of humans. Infrequently, testing the mAb in a species that does not express the epitope or appropriate tissue cross-reactivity profile may be appropriate when the expression of toxicity does not depend on the specificity of the epitope. For example some immunotoxins are known to have primary toxicities that are related to the blood flow of critical organs such as the liver or kidney rather than a consequence of their binding to tumor associated epitope.

TYPES OF NONCLINICAL STUDIES

Pharmacology/Toxicology Safety and Efficacy-related Studies

Non-clinical pharmacology and toxicology studies conducted in laboratory animals and/or *in vitro* systems are integral throughout drug development - from initiation of first-in-human clinical studies through continuing clinical trials and drug approval and at times, in the post-marketing period. The primary goals of non-clinical studies are: to determine a recommend an initial safe starting dose and dose escalation scheme in humans; determine an acceptable risk:benefit

ratio in humans; identify potential target organs of toxicity and/or activity; identify parameters for clinical monitoring; delineate patient inclusion/exclusion criteria; and support labeling claims at the time of approval. The types of non-clinical studies generally expected are *(1) in vivo* and or *in vitro* pharmacology, *(2)* tissue cross-reactivity, *(3)* safety pharmacology, and *(4)* pharmacokinetics and toxicokinetics.

1. In Vivo and In Vitro Pharmacology Studies

Pharmacology studies can be sub-divided into pharmacodynamics and safety pharmacology studies. Nonclinical models are often used to help define mechanisms of action. For cancer therapies, non-clinical pharmacology studies usually consist of both *in vitro* cell culture experiments and *in vivo* animal studies. The specificity of the biological product is determined by performing studies that characterize the *in vitro* binding affinity to the prospective target. In addition, cell culture-based activity assays that are appropriate for the biological product's intended indication and mechanism of action are evaluated. For example a monoclonal antibody could be assessed for antibody-mediated cytotoxicity or anti-angiogenesis activity. Often pharmacology activity assays are the basis for the potency assays required for biologics lot release. In most cases, complementary *in vivo* xenograft studies, in which human cancer cells are implanted into nude or SCID mice, are performed to verify the ability of the biological product to inhibit tumor growth in the context of the normal physiological milieu. The biological product is administered either immediately after cancer cell implantation, to assess antibody influence on tumor implantation and growth, or after tumors are already palpable, to determine whether treatment eradicates established tumors, inhibits further tumor growth and/or impacts survival. Data from such *in vivo* studies can be used to estimate potentially efficacious doses in patients, but also may reveal divergence in pharmacological responsiveness between types of tumor in which stimulation of growth may be observed.

In some cases, if the biological product is intended for use in combination with current cancer therapies or other investigational drugs or biological products, an *in vivo* study is designed to assess the potential influence on known or suspected pharmacological effects.

2. In Vitro Tissue cross-reactivity studies

Data from cross-reactivity studies can serve two purposes. First, data generated using human tissues can verify the anticipated pattern of binding to normal and pathological tissues. Second, tissue cross reactivity studies may identify unintended binding to non-target tissues. Tissue cross-reactivity studies using cryosections of human tissues obtained during surgery or necropsy are an

integral part of the safety assessment of monoclonal antibodies and other biological products. They are often conducted prior to initiating phase 1 trials. Cross-reactivity with non-target tissues is a particular concern with antibodies that mediate antibody- or complement-dependent cytotoxicity (ADCC and CDCC, respectively). Although ADCC and CDCC are therapeutically beneficial for the treatment of cancer, binding to non-target tissue could result in damage in normal tissues. Tissue samples from at least three unrelated human donors should be evaluated for each of the tissues to screen for polymorphism. Positive and negative control tissues and a negative control for the investigational agent should be included in all studies. In the case of a monoclonal antibody, a similar monoclonal antibody that does not bind to the same epitope, i.e., an irrelevant IgG, should be incorporated into the study design. If a conjugated, chemically-modified antibody or an antibody fragment is to be used clinically, it should be tested in that form whenever feasible. Recommendations on conducting these studies are found in the document *Points to Consider in the Manufacture and Testing of Monoclonal Antibody Products for Human Use (1997).*
Comparisons between human tissue cross-reactivity data and tissue cross-reactivity data from different species of laboratory animals aid in the selection of appropriate species for the nonclinical toxicology studies. The selection of a relevant species may also be facilitated by pharmacology studies examining the binding and activity of the investigational agent.

3. Safety Pharmacology Studies

Safety pharmacology studies are a type of specialized toxicology study that investigate the effects of a compound on specified, critical physiological functions, primarily the cardiovascular, respiratory, and central nervous systems (Guidance for Industry S7A Safety Pharmacology Studies for Human Pharmaceuticals, 2001). Safety pharmacology studies are designed to assess the potential for undesirable pharmacodynamic effects following systemic exposure in the therapeutic range and above in specific and critical physiological functions. In the case of highly targeted biologics such as monoclonal antibodies and other biological products, safety pharmacology endpoints can be included in toxicology studies. For certain biological products, additional endpoints might be needed to adequately define the effects of the compound. For example, in the case of anti-CD20 antibodies, flow cytometry can be incorporated into toxicity studies and used to monitor the decline and recovery of $CD20^{+}$ cells.

4. Pharmacokinetic and Toxicokinetic Studies

Pharmacokinetics data can be collected in a separate study or as part of a pharmacology or toxicology study. Kinetics data collected during a toxicology study are referred to as toxicokinetics. Regardless of when they are collected,

kinetics data are very useful in evaluating nonclinical toxicity studies. The information obtained may aid in the design of the clinical trial to avoid toxicities related to systemic levels. Pharmacokinetics studies provide insight into absorption, distribution, and excretion profiles of the clinical product in relevant animal model(s). These studies allow for a more accurate comparison of data obtained in animals to those obtained in humans once the latter become available, and are thus an integral part of safety assessment. Nevertheless, a lack of nonclinical pharmacokinetics data is not likely to preclude the initiation of clinical trials in cancer patients. A lack of pharmacokinetic information may impact on some issues related to a clinical study such as the dose escalation scheme. Formal metabolism studies or mass balance studies are not commonly done for biological products due to the nonspecific degradation pathway for proteins. Similarly biodistribution studies using radiolabeled proteins and other biological products can be difficult to interpret due to recycling of amino acids.

For diagnostic or therapeutic radiolabeled monoclonal antibodies, dosimetry estimates are typically obtained in studies on laboratory animals. The amount of radiation and residence time per tissue are assessed extrapolated to humans with simulation models.

Toxicology studies

Selection of an appropriate animal model is essential for an accurate assessment of non-clinical safety. Toxicology studies conducted in non-relevant species can be misleading and are discouraged. A pharmacologically relevant model expresses an antigenic epitope or receptor and possess similar in activity for the biological product intended for use in humans. A number of techniques, including immunochemical and functional, can be used to identify a relevant species. In the case of monoclonal antibodies, tissue cross-reactivity studies may be conducted using animal tissues to ensure that the animal model selected for toxicology studies exhibits a staining pattern similar to humans. Ideally, toxicology studies should be conducted in 2 species; however, in the case of biological products, the number of species considered appropriate for safety evaluation is strongly influenced by the number of relevant species available and the pharmacology of the product. Frequently, non-human primates are the only relevant model for toxicology studies. In some exceptional cases, biological products are active only in humans and chimpanzees; however, because of their protected status and the experimental limitations associated with them in general, there are many restrictions to conducting non-clinical studies in chimpanzees. When a relevant species cannot be identified for toxicity testing an alternative approach using an analogous protein that recognizes the epitope/receptor in a rodent species or other animal species may be used. While this approach permits conducting the non-clinical toxicity studies, it is not without disadvantages. Chiefly the compound that is being studied may differ pharmacologically and toxicologically in unidentified ways from that being

developed for use in humans. Also with regard to production process, the range of impurities and contaminants may be different when compared to the biological intended for human use. Conducting appropriate studies to define the pharmacology of the analogous protein to the greatest extent possible and to compare it to the product in development can reduce the impact of these disadvantages.

Many biologics intended for human use may induce an immunogenic response (see section below) in laboratory animals. Thus, immunogenicity data from laboratory animals are necessary to accurately interpret a toxicology study, as an immune response to the product can alter its clearance, neutralize its activity, and result in immune complex deposition. However, it is also possible for immunogenicity to have no influence on the toxicology study. Although anti-product antibodies can limit the duration of toxicology studies, their detection should not necessarily result in early termination of the study unless they neutralize activity or enhance elimination. It is important to note that the induction of antibody formation in animals is not necessarily considered predictive of a potential for antibody formation in humans.

General toxicology studies define the toxicity profile of a product and include endpoints that can identify effects on virtually all organ systems. The duration of exposure, frequency of treatment, and route of administration used in general toxicology studies should be as close to the clinical scenario as possible in terms of dose, regimen and duration. In some instances, treatment of clearly responding patients may be continued beyond the duration of treatment achieved in toxicology studies if the risks to the patient are appropriate and the trials are carefully monitored.

The need for specialized toxicology studies that focus on specific systems or functions is based on the nature of a biological product's target, types of effects observed in general toxicology studies and in clinical trials with similar products. Additionally, the stage of disease and patient population may be factors that influence the extent and nature of specialized toxicology studies.

Dosing, Duration and Route of Administration in Toxicology Studies

As a general principle, laboratory animals used in toxicology studies should receive at a minimum the same number of doses as planned for the clinical study. Additionally, the frequency of dosing should be as close as possible to that planned clinically considering the relevant differences in pharmacology and pharmacokinetics across species. Repeated-dose toxicity studies should incorporate toxicokinetics (pharmacokinetic data obtained during a toxicity study), and an assessment of immunogenicity. These data will allow for a more accurate interpretation of toxicology findings and a more meaningful comparison to human data once human pharmacokinetics data are available. For anticancer

biologic agents intended for short-term use (e.g. ≤ 7 days) and for 'acute' life threatening clinical situations, repeat dose toxicology studies up to 2 weeks duration are considered adequate to support clinical studies as well as marketing authorization. For biological products intended for more continued use, studies of longer duration have generally been considered appropriate. In some instances of nonclinical toxicity, the biologic products have exerted pathophysiological effects anticipated to be cumulative, not readily monitorable in the clinic and not predictable by shorter term nonclinical studies. In general nonclinical studies are expected to have a duration in proportion to the proposed duration of the clinical study; nevertheless, in the early phase of clinical study, protocols that incorporate continued clinical dosing are typically supportable by nonclinical studies of 3 months as steady-state effects are commonly achieved in the nonclinical studies and allow for the expression of toxicities that are gradually achieved. In some instances, animal toxicology studies may be shortened in duration relative to clinical study by increasing the dosing frequency. Shorter duration of nonclinical studies may be acceptable for biological products administered in the clinic on a continuing basis if reversibility of toxicity can be readily demonstrated and toxic effects are not shown to be cumulative with repeat administration when considered in the context of the proposed clinical dosing regimen. Robust sets of toxicology data obtained from 3 month toxicity study are supportive of a more aggressive clinical development program as compared to more limited sets of data.

Toxicology studies should also include a recovery period to assess the reversibility of effects and to evaluate the potential for delayed toxicity. Inclusion of a recovery period is especially important when the clinical intention includes multiple dosing over an extended period of time and the pharmacokinetics of the biological products suggests the likelihood of accumulation occurring upon repeated clinical administration. If toxicity is observed at the end of the recovery period, additional studies to characterize toxicity should be considered. Such studies may be conducted concurrent with clinical investigations depending on the nature and extent of the potential toxicities.

Dose selection for nonclinical toxicology studies is another important consideration. A toxicology study should have at least three treatment groups and an appropriate control group. The high dose should produce measurable effects, and the low dose should establish a no-observed-adverse-effect level (NOAEL). In some instances a low dose in animals yielding acceptable toxicity (LoTox) is also acceptable in establishing a starting dose for clinical investigations. A tentative safety factor (NOAEL or LoTox divided by the proposed, initial clinical dose) is an endpoint derived from the animal toxicology study in relationship to the proposed clinical trial. With some notable exception such as with conjugated toxins, biological products often exhibit a high number of safety factors based on animal findings and initial clinical dose. Initial clinical doses are often selected to yield effective systemic pharmacologically levels of the biological product in subjects or patients. Low safety factors based on

doses derived from laboratory studies in animals can be accepted for serious and life-threatening diseases such as cancer.
The route of administration of the biological product used for animal toxicity studies should be the same as that proposed for clinical study. If the human route of administration cannot be duplicated in the animal, then alternative routes may be considered that provide similar exposure to potential target organs of toxicity.

Immunogenicity in Nonclinical Studies

Monoclonal antibodies and therapeutic proteins intended for use in humans can be immunogenic in laboratory animals. Measuring antibodies to these products should be incorporated into all toxicology studies, and their appearance correlated with effects observed in the study. Because immunogenicity in animals is not predictive of a similar response in humans, antibody formation in animals is not generally applied to safety assessment as are data from other endpoints. Rather, antibody formation to the investigational product in animals is used to interpret results of toxicology studies particularly in regards to systemic exposure and some types of toxicity. An antibody response can have no effect on a toxicology study or can impact it through neutralizing the antibody's activity or by changing its pharmacokinetics. Low titers of an antibody response to the investigational product do not necessarily require modification or premature termination of a toxicology study.

Clinical Protocol and its Relationship to the Nonclinical Toxicology Study

In designing nonclinical toxicology studies, it is important to consider the planned clinical protocol (Guidance for Industry M3 Nonclinical Safety Studies for the Conduct of Human Clinical Trials for Pharmaceuticals, 1997). The route of administration, duration of dosing and frequency of administration in the nonclinical toxicology studies should be as close as possible to those planned clinically. Additionally, consideration should be given to the pharmacokinetics and bioavailability in humans compared to animals. If the half-life of a compound is considerably shorter in animals than in humans, increasing the dosing frequency in animals might be indicated.

Compound Comparability and Changes in Formulation

Changes in manufacturing and formulation may occur during the development of biological products. Verifying that a product manufactured using one procedure is comparable to the product or ensuring that a change in formulation is appropriate for use in the clinic is a key aspect of the nonclinical assessment for safety and potential efficacy. Comparability is always evaluated on the basis of

biochemical and biological characterization (i.e. identity, purity, stability, and potency) but in some cases additional nonclinical studies may be required that include pharmacokinetic or pharmacodynamic endpoints.

GLP Compliance

Toxicity studies and safety pharmacology studies are expected to be performed in compliance with Good Laboratory Practice for Nonclinical Laboratory Studies (GLP) (21 CFR 58) in order to ensure the quality and integrity of data generated in nonclinical studies to support the safety of products intended for human use. Some studies employing specialized test systems that are often required for biopharmaceuticals may not be able to comply fully with GLP. Areas of non-compliance and deviations are identified and their significance is evaluated relative to the impact on the overall safety assessment.

Pre-Clinical Requirements for Phase 1 Studies – Biological Oncology Products for March 13, 2006 meeting of ODAC

The review and regulatory decision making concerning biological oncology products has undergone and continuous to undergo a series of important changes. These changes include both administrative and scientific aspects. Of particular concern is the relationship of the duration of nonclinical studies needed to provide sufficient information to the proposed length of dosing in clinical studies in terms of the initial period of clinical study and the continuation of patients subsequently. To address this point, a continuous series of INDs submitted between July 2001 through November 2005 and that were the subjects of a new molecular entities (NMEs) were entered into a database. To be included in the database, the NMEs were proposed for study as therapeutic agents but were not approved biological products. Additionally, INDs were not included in the database if the investigational agent was intended for study as a diagnostic, use as a radiolabeled compound, supportive therapy, single patient or emergency use, Fifty-one NMEs comprised the database containing monoclonal antibodies (72.5%), fusion proteins (15.7%), cytokines (3.9%) and other (7.8%). A number of endpoint relative to the types and nature of nonclinical toxicology studies were examined as submitted in the INDs. The duration of toxicity studies submitted in the IND ranged from ≤ 1week to > 3 months. In 4% of the INDs no toxicity study was performed. In terms of the number of days of dosing and dosing frequency, toxicology studies conducted in laboratory animals were in general agreement and matched or often exceeded the proposed clinical protocol. The frequency and basis for clinical holds decision and recommendations was also reviewed. In the vast majority of cases, potential clinical holds were resolved prior to the end of the 30 day period of review. When clinical holds involved pharmacology and toxicology safety concerns, they often occurred in conjunction with other safety issues related to medical or CMC

aspects of the IND. Lack of adequate duration in nonclinical studies relative to the proposed length of the clinical protocol was an important factor in some clinical hold decisions. In these cases a longer duration of nonclinical study was considered important to the assessment of safety and additional nonclinical studies were requested. Specific examples will be presented of emergent toxicities that were observed or toxicities that significantly intensified over the period of time between 4 weeks and 3 months.

Cancer Chemother Pharmacol (1998) 41: 173–185

ORIGINAL ARTICLE

Joseph J. DeGeorge · Chang-Ho Ahn
Paul A. Andrews · Margaret E. Brower
Diana W. Giorgio · M. Anwar Goheer
Doo Y. Lee-Ham · W. David McGuinn
Wendelyn Schmidt · C. Joseph Sun
Satish C. Tripathi

Regulatory considerations for preclinical development of anticancer drugs

Received: 19 December 1996 / Accepted: 2 June 1997

Abstract The entry of new anticancer treatments into phase I clinical trials is ordinarily based on relatively modest preclinical data. This report defines the battery of preclinical tests important for assessing safety under an Investigational New Drug application (IND) and outlines a basis for extrapolating starting doses of investigational anticancer drugs in phase I clinical trials from animal toxicity studies. Types of preclinical studies for the support of marketing of a new anticancer drug are also discussed. This report addresses differences and similarities in the preclinical development of cytotoxic drugs (including photosensitizers and targeted delivery products), drugs used chronically (chemopreventive drugs, hormonal drugs, immunomodulators), and drugs intended to enhance the efficacy (MDR-reversing agents and radiation/chemotherapy sensitizers) or diminish the toxicity of currently used anticancer therapies. Factors to consider in the design of preclinical studies of combination therapies, alternative therapies, and adjuvant therapies in the treatment of cancer, and to support changes in clinical formulations or route of administration, are also discussed.

Key words Antineoplastic agents · Toxicity tests · Toxicology · Guidelines · Phase I clinical trials

This article is not an official FDA guidance or policy statement. No official support or endorsement by the Food and Drug Administration is intended or should be inferred

J.J. DeGeorge · C-H. Ahn · P.A. Andrews (✉)
M.E. Brower · D.W. Giorgio · M.A. Goheer ·
D.Y. Lee-Ham · W.D. McGuinn · W. Schmidt ·
C.J. Sun · S.C. Tripathi
Division of Oncology Drug Products,
Center for Drug Evaluation and Research,
Food and Drug Administration, 5600 Fishers Lane HFD-150,
Rockville, Maryland 20857, USA
Tel. (301)-594-5782; Fax: (301)-594-0499;
E.mail: andrewsp@cder.fda.gov

Introduction

Malignant, nonresectable cancers are life-threatening, and aggressive measures are used in treating them. Antineoplastic therapies frequently include toxic chemicals or biological products that are designed to destroy tumor tissue or halt cell replication. Despite the serious toxicities of many anticancer drugs, careful dosing, clinical monitoring and prompt treatment of toxicity makes the side effects less threatening to a patient than their disease. Since it is recognized that doses of anticancer drugs high enough to kill cancer cells usually induce serious side effects in patients, the preclinical testing of oncology drugs differs from testing of nononcology drugs. The Division of Oncology Drug Products within the Center for Drug Evaluation and Research (CDER) at the US Food and Drug Administration (FDA) recognizes the urgency of development of new anticancer drugs and the need to rapidly move promising agents into clinical studies. This report offers a regulatory perspective on the preclinical development of new anticancer drugs that is intended to clarify the differences from the preclinical testing of nononcology drugs and to describe the data that are important to support human testing and eventual marketing.

The types of preclinical studies expected for support of clinical trials and then marketing of a new drug depend on both the intended use of the drug and the population of patients being studied and treated. In situations where potential benefits are greatest (advanced, life-threatening disease), greater risks of treatment toxicity can be accepted and the required preclinical testing can be minimal. In cases where the patient population is free of known disease (e.g. adjuvant therapy or chemoprevention) the acceptable risks are much less and preclinical evaluation should be more extensive [32]. The toxicities of many modulating agents intended to enhance the efficacy or diminish the toxicity of anticancer agents are more similar to those of

nononcology therapies. However, these modulating agents could enhance the toxicity or diminish the activity of cytotoxic drugs by altering their toxicodynamics, pharmacodynamics, and pharmacokinetics. Thus, toxicological evaluation in combination with the modulated cytotoxic drug is an important part of preclinical development.

The following considerations are offered in an effort to balance the risks to be borne by the proposed patient population and the realities of drug testing in humans. The differences in preclinical testing between cytotoxic, chronic (i.e. adjuvant therapy, chemopreventive drugs, hormonal drugs, and immunomodulators), and modulating therapies are emphasized. Issues of chemistry and manufacturing controls, clinical study design, and development of biologic agents for cancer treatment are beyond the scope of this report. If the appropriate preclinical development strategy remains uncertain after contemplating the following considerations, then sponsors are encouraged to initiate pre-IND discussions with Division staff regarding their preclinical study plan.

General considerations for anticancer drug development

Preclinical studies of anticancer agents

The safety of first-time use in humans is assessed through preclinical studies of pharmacodynamics, pharmacokinetics (toxicokinetics), toxicity, and their relationships. The purposes of these safety studies are: (a) to determine a starting dose for clinical trials that is both reasonably safe and allows for possible clinical benefit for the patient, (b) to identify potential end-organ toxicities and determine their reversibility, and (c) to assist in the design of human dosing regimens and escalation schemes for clinical trials. Animal toxicity studies most effectively accomplish these objectives when performed using schedules, durations, formulations, and routes comparable to those proposed in clinical studies. Use of longer duration preclinical studies may lead to underestimates of the appropriate clinical dose, while shorter studies may not identify cumulative dosing toxicities. The toxicity studies should generally conform to the protocols recommended by the National Cancer Institute for toxicology assessment for anticancer agents[1] and are expected to be conducted in accordance with Good Laboratory Practices (GLP) [16, 17]. When studies are not performed according to GLP, deviations should be documented and the potential impact of these deviations on study outcome and credibility should be described [16, 17].

Typically, only two toxicology studies are essential to support initial phase I clinical trials in patients with advanced cancers (Table 1). The first of these is usually a study in rodents that identifies doses that produce life-threatening and non-life-threatening toxicity. The second study should determine whether doses identified as tolerable in rodents produce life-threatening toxicity in a non-rodent species. At least one of these studies should assess clinical signs, body weight, food consumption, clinical pathology, and gross pathology over a range of doses from nontoxic to toxic and should include an examination of histopathology at doses that cause toxicity (or at the highest dose tested). Genotoxicity tests are not generally needed for cancer chemotherapies to support testing in phase I clinical studies unless healthy volunteers will be entered into the study.

While not essential, information on the pharmacodynamics and pharmacokinetics of drugs is extremely valuable for supporting the safety profile and can significantly contribute to the efficiency of drug development. A phase I study may be conducted with no in vitro or in vivo preclinical pharmacodynamic information, but preclinical studies on biological activity and efficacy can substantially aid in clinical study design. Such studies help estimate effective dosages, dosing schedules, and optimal plasma concentrations. This information is likely to be particularly useful when developing noncytotoxic agents. It may be desirable to develop such agents (e.g. immunomodulators) by escalating the human dose to a pharmacodynamically active range rather than to the maximum tolerated dose (MTD). Pharmacokinetic data can be gathered as a part of pharmacology or toxicity studies and do not usually need to be collected separately. Single- and multiple-dose pharmacokinetic studies in the most appropriate species are best performed using dosing schedules, durations, and routes comparable to those that will be used in clinical studies [15]. The pharmacokinetic information obtained assists the evaluation of animal toxicity and efficacy, and may suggest modifications in the intended dose, route or schedule for the clinical trial. The importance of the parameters being measured will vary depending on the clinical trial design and therapeutic classes as discussed in the subsections below. In combination with pharmacodynamic data, this information can be used to help calculate initial doses in humans that have a greater likelihood of activity without adversely affecting safety, and can contribute to optimal dose escalation in early clinical studies.

The proposed therapeutic indication, the outcome of early clinical development, the nature of toxicities seen in animals and in humans, and the projected duration of clinical treatment all determine the preclinical studies necessary to support a New Drug Application (NDA). In general, for oncology drugs, sponsors should conduct toxicity studies using the same schedule and duration of administration as the intended clinical treatment cycle (Tables 1–3). Cytotoxic drugs used to treat advanced disease rarely need studies with more than 28 days of dosing submitted with the NDA (Table 1). In contrast, for drugs intended for continuous

[1] The Developmental Therapeutics Program; Division of Cancer Treatment, Diagnosis, and Centers; National Cancer Institute (Rockville, MD USA) may be contacted for protocol details

Table 1 Preclinical studies for cytotoxic oncology drugs

Stage	Category	Issues to be addressed	Studies considered important[a]	Studies considered useful
IND	All cytotoxics	Starting dose, end-organ toxicities	Rodent[b] and nonrodent[c] toxicology[d]	
		Genetic toxicity	Genetic toxicity panel[e]	
		Effective concentrations, schedule		Pharmacokinetics, pharmacodynamics
	Modifications for Special Categories			
	Photosensitizer	Systemic toxicity	Toxicology studies in subdued light	
		Phototoxicity		In vivo study with illuminated skin
		Plasma $t_{1/2}$		Pharmacokinetics
	Antibody conjugate	Stability	Stability in plasma	
		Toxicity of drug alone	Toxicology in one species	
		Specificity	Human tissue screen	Activity in cell lines ± target antigen
		Pharmacokinetics		Pharmacokinetics
	Liposomal delivery	Drug product toxicity	Include free drug and blank liposomes in toxicity testing	
		Pharmacokinetics versus free drug	Pharmacokinetics	
	Depots	Drug product toxicity	Include free drug and empty depot in toxicity testing	
		Toxicity to contacted tissues	Histopathology of depot site	
NDA	All cytotoxics		Rodent and nonrodent toxicology[a,g], genetic toxicity, stage C-D teratogenicity[f] in rodents and nonrodents	Targeted special toxicity

[a] In general, the schedule and duration of administration in the toxicology study should mimic the clinical trial
[b] Should determine the dose severely toxic to 10% of the animals (STD_{10})
[c] Should determine toxicity of one-tenth the rodent STD_{10} on a mg/m^2 basis
[d] One study should include histopathology
[e] Only for phase I testing in normal volunteers or patients believed to be disease-free
[f] Should be submitted during development
[g] Studies with more than 28 days of dosing are rarely needed

daily administration such as for chemoprevention, adjuvant therapy, or long-term hormonal or immunomodulation therapy, chronic studies should be conducted up to a maximum of 6 months in rodent and 12 months in nonrodent species (Table 2). International Conference on Harmonization (ICH) stage C-D[2] reproductive toxicity studies in a rodent and a non-rodent species are important components of the preclinical evaluation of anticancer drugs and should be submitted early in development [14].

Carcinogenicity studies are not required for cytotoxic drugs used to treat advanced systemic disease, but can be important in the assessment of drugs intended for chronic use for chemoprevention, adjuvant, or hormonal therapy when patients are likely to have a long survival [18]. The current standard is the 2-year rodent bioassay [47], although alternatives may be suitable [20]. Depending upon the nature of toxicities seen with the drug or drug class in animals and in humans, targeted special toxicity studies to support NDA filing may also be needed. For example, in the development of anthracyclines and platinum drugs, which are known to have cardiotoxic and ototoxic potential, respectively, additional preclinical cardiotoxicity and ototoxicity studies have been useful [11, 28, 36, 46]. In addition, neonatal reproductive toxicology and DNA adducting studies have been useful in the development of antiestrogenic agents [5, 30, 31, 37, 44, 45]. A discussion with FDA staff on the preclinical studies needed for marketing approval for a particular drug is recommended at or before the end of phase II clinical studies.

[2] ICH stage A-B, C-D, and C-F reproduction toxicity studies correspond to the previously designated segment I, II, and III studies which are defined by daily administration of drug, respectively, during the period from premating to implantation, implantation to birth (period of organogenesis), and implantation to sexual maturity [14]

Table 2 Preclinical studies for noncytotoxic, chronically administered oncology drugs

Stage	Category	Studies considered important	Studies considered useful
IND	All noncytotoxic chronic therapy	Rodent[a] and nonrodent[b] toxicology[c,d]	Pharmacokinetics, -dynamics
	Modifications for special categories		
	Adjuvant therapy	Genetic toxicity panel[e]	
	Chemopreventive	Toxicology studies should also define NOAEL, Genetic toxicity panel	Efficacy studies Carcinogenicity[e] Stage A-B reproductive toxicity Stage C-D teratogenicity[e]
	Hormonal	28-day toxicology studies usually suffice for limited phase I/II testing in advanced cancer, genetic toxicity panel[e]	
	Immunomodulator	28-day toxicology studies usually suffice for limited phase I/II testing in advanced cancer, genetic toxicity panel[e], define dose versus immunologic response curve to identify shape (bell-shaped?) and surrogate markers	
NDA	All non-cytotoxic chronic therapy	Toxicology studies of equivalent duration to labeled use up to 6 months in rodents and 12 months in nonrodents, genetic toxicity panel, carcinogenicity[f], stage C-D teratogenicity in rodents and non-rodents	
	Additional for hormonal	Stage A-B reproductive toxicity	Stage C-F reproductive toxicity, neonatal reproductive tract toxicity, DNA adducting (drug specific)
	Additional for chemopreventive	Stage A-B and C-F reproductive toxicity carcinogenicity (always)	

[a] Should determine the dose severely toxic to 10% of the animals (STD_{10})
[b] Should determine toxicity of one-tenth the rodent STD_{10} on a mg/m^2 basis
[c] In general, the schedule of administration in the toxicology study should mimic the clinical trial with a duration as long as the intended clinical study up to 6 months in rodents and 12 months in non-rodents
[d] One study should include histopathology
[e] Expected prior to clinical testing in patients with low risk of cancer recurrence, or testing in healthy volunteers
[f] May be unnecessary depending on intended patient population [18]

Starting doses and dose escalation

As described above, one of the primary goals of preclinical studies is to estimate a safe starting dose for the initiation of phase I trials in humans. The starting dose for clinical trials with cytotoxic drugs for oncology indications has traditionally been one-tenth the dose lethal to 10% of rodents on a body surface area basis (milligrams per meter squared) [23, 29, 35]. Studies that actually measure death as an endpoint, however, are not required so long as the dose range studied includes doses that cause severe, life-threatening toxicity. Thus, the starting dose is generally now chosen as one-tenth of the dose that causes severe toxicity (or death) in 10% of the rodents (STD_{10}) on a milligrams per meter squared basis, provided that this starting dose, i.e. one-tenth the STD_{10}, does not cause serious irreversible toxicity in a nonrodent species [29, 35]. If irreversible toxicities are produced at the proposed starting dose in nonrodents (usually dogs) or if the nonrodent is known to be the more appropriate animal model, then the starting dose would generally be one-sixth of the highest dose tested in nonrodents that does not cause severe, irreversible toxicity[3]. In some cases, rodents or dogs may not be appropriate species because they do not model the relevant human biochemical or metabolic processes. For example, folate pools in rodents greatly exceed those in humans [4], so that rodents are generally inappropriate species for testing antifolates. Also, dogs poorly predict the toxicity of some platinum analogues, and an alternate animal model might be preferred [34]. Knowledge of relevant physiological, biochemical, and pharmacokinetic differences between humans and animal models can help determine the most appropriate species to be used for selecting a starting dose. Whenever feasible, these starting doses should be

[3] This calculation is the same as taking one-third of the toxic dose low (TDL) [29, 35]. We believe the current expression of "one-sixth the highest non-severely toxic dose" is simpler and can be applied to the data more universally than taking, in practice, "one-third the dose which causes toxicity but when doubled does not kill the nonrodents". Frequently, the TDL cannot be technically defined in many studies

Table 3 Preclinical studies for modulators of oncology drugs

Stage	Category	Issues to be addressed	Studies considered important[a]	Studies considered useful
IND	All modulators	Starting dose, end-organ toxicities	Rodent[b] and non-rodent[c] toxicology[d]	
		Genetic toxicity	genetic toxicity panel[e]	
		Effective concentrations, schedule		Pharmacokinetics
	Additional studies for special categories			
	MDR modulator	Combination toxicity	One species at minimally and significantly toxic doses of cytotoxic	In vivo efficacy of combination
		Pharmacokinetic perturbations	Pharmacokinetics	
	Chemosensitizer	Combination toxicity	One species at minimally and significantly toxic doses of cytotoxic	
	Radiation sensitizer	Delayed toxicity to normal tissues		Skin/leg contracture
	Chemoprotection	Combination toxicity, tumor protection	In vivo efficacy of combination with histopathology	
NDA	All modulators		Toxicology studies of equivalent duration to labeled use up to 6 months in rodents and 12 months in non-rodents, genetic toxicity, stage C-D teratogenicity in rodents and non-rodents	Targeted special studies

[a] In general, the schedule and duration of administration in the toxicology study should mimic the clinical trial
[b] Should determine the dose severely toxic to 10% of the animals (STD_{10})
[c] Should determine toxicity of one-tenth the rodent STD_{10} on a mg/m2 basis
[d] One study should include histopathology
[e] Only for phase I testing in normal volunteers or patients believed to be disease-free

calculated from studies using the proposed clinical route, schedule, and duration.

The dose escalation scheme for phase I clinical studies often follows the standard or modified Fibonacci procedure [10]. Examples of other common and acceptable approaches include modified continual reassessment methods [13, 39] and pharmacokinetically guided dose escalation strategies [8]. These alternatives often necessitate a more extensive preclinical evaluation. For example, pharmacokinetic guidance of dose escalation is most effectively applied when: (a) linear pharmacokinetics are observed at drug concentrations spanning the pharmacological and toxicological effects, (b) the area under the drug concentration versus time curve (AUC) at the mouse STD_{10} can be defined, (c) protein binding in mouse and human plasma has been quantified, and (d) it is known whether metabolites contribute to the toxic effects [7, 8, 27, 40]. Although preclinical studies are used to determine the starting dose for phase I clinical trials, the highest doses for oncology drugs are rarely restricted by the doses used in preclinical toxicology studies as long as the toxicities of the new anticancer drug can be readily monitored, are reversible, and sufficiently precede lethality in animals. Instead, the maximum dose is restricted by the toxicity observed in the clinical trial, judged most often using NCI/DCTDC Common Toxicity Criteria [38].

Considerations for specific cytotoxic therapies

Combinations of cytotoxic agents

The evaluation of cytotoxic agent combinations has traditionally been conducted in the clinical setting using an empirical approach. This has generally been successful, but may not be optimal. Preclinical studies provide an opportunity to explore a variety of doses, dose ratios, and schedules to optimize benefit and minimize toxicity. Nonetheless, unless there is reason to believe that synergistic interactions occur that would substantially increase the toxicity of the combination, preclinical testing is not considered essential provided that each agent has been fully evaluated in humans. When synergistic effects may be anticipated such as when one agent interferes with the metabolism or elimination of the other agent or both cytotoxic agents target the same metabolic pathway or cellular function, preclinical testing of the combination is desirable.

Photosensitizers

One class of cancer chemotherapeutic drugs is therapeutically inactive until irradiated with light. These

photosensitizers or phototherapy agents usually form radicals after absorbing light energy that are ultimately responsible for tumor destruction. In photosensitizer therapy, tumor tissues are typically irradiated with laser light. When there is a choice, longer wavelengths of the irradiating light are preferred because they cause less direct tissue damage and because they penetrate more deeply into tumor tissue than shorter wavelengths.

Selective damage to tumor tissue is obtained by directing the activating light to the tumor. In addition, most phototherapy compounds concentrate in tumor tissues more than in surrounding normal tissue when given systemically. This increased concentration of photosensitizer combined with localized irradiation can kill tumor cells with great selectivity. Nevertheless, when these compounds are given systemically they commonly distribute in appreciable concentrations in all tissues and this provides the potential for toxicity. When these drugs accumulate in the eye or skin, patients may suffer irreversible retinal damage or severe phototoxicity similar to sunburn when exposed to ambient light [12]. Thus, it is important to know the plasma elimination half-life (and, if possible, tissue elimination half-lives) in preclinical studies so that the length of time a patient should protect themselves from light can be estimated.

Standard toxicity studies with multiple dose levels should be conducted in subdued illumination to clearly define the systemic toxicities of the photosensitizer. Subdued lighting allows systemic toxicities to be more clearly distinguished from phototoxicities. In addition to these standard toxicity studies, it is beneficial to assess phototoxicity before phase I clinical investigation begins because these drugs can cause prolonged photosensitivity. Acceptable models for these photosensitivity tests are either hairless or appropriately shaved species. The photosensitivity assessment should include toxicity testing as a function of both light dose (total energy) and drug dose and should ideally determine the duration of sensitivity in relation to plasma levels of the photosensitizer. Since a primary concern for the patient is the toxicity related to sunlight exposure, the light source for these tests should have a spectral distribution that approximates sunlight. Frequently, doses that are well below the no observable adverse effect limit (NOAEL) when the animal is housed in subdued light are lethal when the animal is briefly irradiated. Even though the photodynamic effect is expected to affect only tissues that are exposed to the light source, there is concern that photodegradation products could cause distant toxicities. Therefore, these phototoxicity tests usually include standard assessments of clinical signs, clinical pathology, gross pathology, histopathology of major organs, and the reversibility of toxicities. Clinical photodynamic therapy does not routinely involve repeated doses, and thus preclinical studies using daily irradiation during repeat dose testing may not be relevant to clinical safety concerns.

Without light these photosensitizers may not cause genotoxicity in standard tests, but subsequent irradiation may cause considerable damage to the DNA of cells exposed to the compound. Thus, genotoxicity tests are best done with and without light. The assessment of clastogenicity and mutagenicity should be done with increasing compound concentrations at a high light dose, and with increasing light dose (total energy) using broad-spectrum light at high compound concentration. The highest doses of drug of each series of tests should be consistent with international standards [19, 22].

In many cases an effective dose of drugs in this class is nontoxic in subdued light and the starting dose can be chosen based on efficacy studies rather than toxicity studies. This pertains only if the projected efficacious starting dose is lower than the safe dose estimated from the toxicity studies.

Specialized drug delivery

Administration of anticancer drugs as depots, attached to carriers, or in specialized encapsulated forms has the potential for significantly improving efficacy. Advantages of specialized drug delivery may include: (a) specific targeting of the drug to the tumor, (b) minimization of toxic side effects, (c) prolongation of therapeutic drug concentrations, (d) improved delivery of hydrophilic drugs to tumor cytoplasm, and (e) practical administration of very lipophilic drugs. Examples of delivery systems include copolymer implants, human albumin microspheres, monoclonal antibody–drug conjugates, and liposomal encapsulation. Development of anticancer drugs administered via carriers or in depots may necessitate additional preclinical evaluation beyond that of conventional cytotoxic drugs.

For antibody–drug conjugates, the two main safety concerns are the potential for toxicity from abrupt release of the drug and the potential for the antibody–drug conjugate to cause unexpected, specific toxicity in normal human tissues. Studies of the stability of the conjugate in human plasma as a function of the proposed release mechanism (e.g. pH if hydrolytic, glutathione concentration if reductive) help determine the necessity of conducting additional toxicology studies [21]. When additional studies are indicated, using the form of the drug released from the conjugate (i.e. including linker groups) may identify clinically important toxicities. Testing the reactivity of the conjugate with a complete panel of human tissues from at least three different sources is suggested [21]. When the target antigen is not expressed in the tissues of the standard preclinical animal models, a tolerance study in Pongidae apes at a dose that is at least double the planned human starting dose should also be considered. Both the reactivity screen and the tolerance study may reveal sites of potential tissue-specific toxicity, while the standard toxicology studies may define nonspecific toxicities. Specificity studies of binding or cytotoxicity in cell lines with and without an expressed target antigen also help to assess whether there is a significant differential between the toxicity to a targeted and nontargeted tissue. If feasible,

pharmacokinetic studies that distinguish between conjugate, free antibody, and free drug are also highly desirable for interpreting toxicology findings and supporting interspecies comparisons. Selection of a starting dose for clinical study should consider not only the results of the toxicity studies with the conjugate, but also the stability of the conjugate and the potential toxicity of released drug.

With liposomal drugs, standard preclinical toxicology studies of the delivery system, free drug, and the final formulation are important for evaluating a drug product's potential for toxicity. Liposomal formulations usually dramatically prolong systemic exposure. Thus, when repeated doses are to be used clinically, it is especially important to study a similar schedule preclinically because of the potential for drug accumulation. When the delivery system is designed to affect drug absorption, distribution, biotransformation, excretion or target organ accumulation, small changes in the design of the delivery system may have substantial effects on overall toxicity. Conducting the toxicity studies with the final formulation can avoid concerns about such effects. Comparative pharmacokinetic studies of the final formulation versus free drug can be very helpful in suggesting schedules and interpreting changes in the spectrum and severity of toxicities. Occasionally, studies of the empty liposomes plus free drug in combination may also be useful for understanding alterations in efficacy seen with the liposomal preparation. For example, blank liposomes may alter the pharmacokinetics of the free drug in a fashion sufficient for therapeutic gain [33].

Preclinical development of depot formulations generally follows that of liposomal formulations. Additionally, a study of the toxicity of the depot in the tissue or compartment intended to be used clinically should be conducted which includes a histopathologic examination of the adjacent tissues. Initial clinical doses similar to the total dose of the drug previously investigated in humans may be used in the absence of significant changes in toxicity profile for the depot formulation.

Alternative therapies

"Alternative" therapies include both single agents and multicomponent entities derived from plants or animals. Herbal products and tissue or fluid extracts from animal sources intended for the treatment or prevention of cancer or precancerous conditions belong in this category. The identity of the active ingredient of these entities is frequently uncertain. Consistency in taxonomic identification, collection, storage, and processing may pose additional difficulties. A useful initial step is to prepare a batch of the drug product large enough to be sufficient for both initial preclinical and clinical studies. The usual battery of toxicology studies for anticancer agents should be conducted unless there is adequate human safety experience. Since it is difficult to correlate specific drug product components with pharmacologic action, attempts should be made early in the development scheme to control the manufacturing processes to produce consistent batches for subsequent preclinical and clinical study. Further efforts should be made in the later stages of development to identify biologic assays which can be used to assure activity and as release specifications for the marketed product.

Herbal products represent a specialized subset of alternative therapies, as there is often significant human experience with their use. If there is a documented history of use of the herbals or if these preparations are freely marketed in the United States, then no preclinical pharmacology or toxicology is required for initial trials using the marketed product. Submission of data on the traditional use, preparation of the product, and safety profile of any known components of the herbal preparation for the IND is encouraged. When a product different from the marketed version is intended for the clinical trial, information on the preparation of the product to be tested is important in determining whether toxicology studies are necessary. If a herbal product is prepared in a manner different from the marketed product (e.g. alcoholic extraction instead of an aqueous preparation such as tea) or administered by an alternative route, then the standard toxicology studies for an investigational anticancer drug may be necessary. As the development of the herbal therapeutic agent continues in expanded trials, animal data including the histopathology, serum chemistry, hematology, reproductive, and genetic effects of the compound should be obtained either through literature data on the individual components of the herbal product or through toxicologic testing.

Considerations for chronic therapies

Chemopreventives

The preclinical development of chemopreventives has been previously described and should proceed similarly to most nononcology drugs [32]. The key considerations are summarized in Table 2.

Adjuvant therapy

The preclinical studies expected for drugs developed for adjuvant therapy depend on the prior human experience with the drug, the anticipated risks and benefits for the intended patients, and the expected mechanism of action. Few drugs are initially tested in humans in the adjuvant setting. Substantial clinical experience with these drugs is thus usual by the time they are considered for therapy in patients who have had their primary tumor removed or controlled. Nonetheless, further preclinical testing may be needed, depending on whether there are changes in the pattern of clinical use.

Additional preclinical studies that focus on long-term toxicity should be conducted for agents with which there is limited long-term clinical experience and intended for chronic treatment of patients in whom the risk of recurrence of cancer is relatively low. Cytotoxic drugs normally do not need additional long-term studies to support adjuvant use because the clinical experience with these drugs is usually extensive, they are usually administered using intermittent cycles rather than daily dosing, and the risks to patients are already well understood. When conducted, long-term studies should use the intended adjuvant route and schedule for at least as long as the intended clinical treatment duration, up to a maximal duration of 6 months in rodents and 12 months in nonrodents (usually dogs). A complete battery of genetic toxicity tests should be conducted prior to trials in patients believed to be free of disease. Carcinogenicity studies are usually expected prior to application for market approval.

Hormonal drugs

The mechanism of action of hormonal drugs differs significantly from that of other antineoplastic agents. These drugs are usually not directly cytotoxic, but may act as antiestrogens, progestins, antiprogestins, androgens, antiandrogens, aromatase inhibitors, or gonadotropin releasing hormone agonists. As with cytotoxic therapies, the preclinical toxicity assessment of hormonal drugs should use a similar route, schedule, duration of treatment, and formulation of drug substance as that proposed in clinical therapy. Standard 28-day toxicology studies with daily drug administration usually support small, phase I and phase II clinical trials with advanced-stage cancer patients. As clinical studies with longer durations of treatment are planned in patients likely to have an extended survival, additional preclinical testing usually follows the standard practice that the duration of the toxicology study be at least as long as the clinical trial. Because hormonal agents are generally used over an extended period, the complete toxicology assessment may need to focus on long-term effects on organ systems. Maximal duration of treatment in animals is usually limited to 6 months in rodents or 12 months in nonrodents. Although these agents are customarily developed for sex-specific indications, preclinical testing of both sexes allows identification of toxicities unrelated to the primary hormonal action of the drug that may be obscured in animals of the same sex as the intended treatment population. In addition, sex-based differences in the nonreproductive organ toxicities, sensitivity, or metabolism of a given drug may not be correlated across species [6]. Testing of hormonal agents in both sexes is thus more likely to provide the full spectrum of potential toxicities associated with a drug's use.

It is expected that the standard battery of genotoxicity tests assessing mutagenicity and clastogenicity will be conducted prior to phase I testing in patients believed to be disease-free [19, 22]. Carcinogenicity studies are expected if the hormonal drug is intended for use in patients believed to be disease-free or as adjuvant therapy. Studies should be conducted to evaluate reproductive performance and fertility in rats (ICH stage A–B, segment I), and teratogenicity in rats and rabbits (ICH stage C–D, segment II). Depending on the patient population and the duration of hormonal therapy, ICH stage C–F (segment III) studies may be needed. Many estrogen agonists or antagonists are structurally or pharmacodynamically related to diethylstilbestrol (DES), which is known to cause reproductive tract malignancy and abnormalities in humans exposed in utero [25]. Testing the potential of compounds related to DES to cause reproductive tract changes in neonates and pubescent animals is therefore considered important [5, 31, 44, 45]. Such studies typically focus on reproductive tract development following 3–5 days of dosing in rodent neonates in order to observe such pathologies as vaginal adenosis [44, 45].

Immunomodulators

Therapeutic agents that modulate the body's immune response to cancerous cells usually do so at concentrations significantly lower than those that cause severe toxicities in animals of the type seen with standard cytotoxic agents. Some biological responses to immunomodulators are species specific and may be related to toxicity to the immune system. Non-species-specific toxicities, however, do occur that are not directly related to modulation of the immune system. Thus, a standard safety evaluation conducted to identify these non-species-specific toxicities is important. As for hormonal agents, standard 28-day toxicology studies with daily drug administration are adequate to support initiation of phase I and phase II clinical trials that enroll advanced-stage cancer patients. In addition to the toxicology studies, knowledge of the mechanism of action also contributes to the evaluation of the safety of immunomodulators and selection of a starting dose. Studies that combine a measurement of the appropriate immunological response in addition to toxicity assessments are particularly useful because these agents, unlike most other drugs used to treat cancer, have sometimes exhibited bell-shaped dose response curves for desired activities. It is therefore especially important to use a starting dose that does not exceed the beneficial therapeutic range. The low doses often administered early in the phase I study sometimes give plasma concentrations of immunomodulators that preclude conventional pharmacokinetic study. In lieu of pharmacokinetic data, it may be useful to provide animal data on possible surrogate endpoints of activity that can be used in the clinic to demonstrate that active concentrations have been reached. Surrogate markers of activity that have been assessed include induction of interferon,TNF-α, neopterin, or β–2-microglobulin [41, 49].

Considerations for modulating therapies

Multidrug resistance-reversing agents

Prior to therapy, during therapy, or at the time of relapse, many tumors develop resistance to a variety of structurally unrelated anticancer drugs. This phenomenon is termed multidrug resistance (MDR). Mechanisms of MDR include, but are not limited to, altered expression of P-glycoprotein (P-gp), MDR-associated proteins (e.g. MRP and LRP), topoisomerases, and glutathione-*S*-transferases. Currently most MDR-reversing agents under development target the P-gp-dependent mechanism. P-gp is encoded by the *mdr*1 gene that is often amplified or overexpressed in MDR-manifesting tumors [26]. By functioning as an efflux pump, P-gp causes decreased drug accumulation and reduced cytotoxicity of anticancer drugs in tumor cells. P-gp is also expressed in many normal tissues (e.g. in the gastrointestinal tract, brain, kidney, and liver) [9]. One role of P-gp expression is presumably to carry out the efflux of toxic substances from these tissues. The inhibition of the efflux function of P-gp by a MDR-reversing agent increases intracellular concentrations of the cytotoxic drug in tumor tissue expressing P-gp. However, inhibition may also increase levels of cytotoxic drugs in normal P-gp-expressing tissues, potentially resulting in alterations of the severity and types of toxicities usually associated with the cytotoxic drug alone [1]. Furthermore, clinical and preclinical studies have shown that drugs interacting with P-gp can significantly alter the pharmacokinetics of cytotoxic drugs [2].

In view of the added risks associated with the combination of a MDR-reversing agent and a cytotoxic drug(s), the following preclinical studies are considered important for determining the safety of a proposed clinical trial. First, a standard profile of toxicology studies for the MDR-reversing agent alone should be conducted which take into consideration the likely duration of use in early clinical trials. Second, a study of the MDR-reversing agent combined with the cytotoxic drug in one species (usually a rodent) should be conducted to assess toxicity at both minimally and significantly toxic doses of the cytotoxic agent (Table 3). This information may also be derived from in vivo combination efficacy studies when an assessment of toxicity has been included. Based on experience to date, a combination study with one cytotoxic drug from a structurally related therapeutic class generally suffices for determining the safety of the modulator with all cytotoxic drugs in that class. Third, appropriate pharmacokinetic parameters should be derived since pharmacokinetic changes have often been shown to be important in interpreting the toxicity from such combinations.

There are several approaches for the selection of starting doses and escalation schemes for combinations of anticancer drugs and MDR-reversing agents. Some investigators have chosen to use a relatively high dose (or effective concentration) of MDR reverser and to escalate the anticancer drug. Others have started with a relatively high dose of the cytotoxic drug and escalated the MDR reverser. Neither of these approaches has been established as superior. Preclinical studies can guide either approach to dose selection by establishing a ratio of toxicity or potential toxicity of a given dose of the cytotoxic drug in the presence and absence of the MDR reverser. Acceptable endpoints for establishing this ratio might include direct measures of severe toxicity, such as marrow suppression or lethality, or measures of plasma concentrations. For example, when a therapeutic dose of a reversing agent increases the AUC of the cytotoxic drug fivefold in a preclinical model, a starting dose of the cytotoxic drug that is decreased by a factor of five from the accepted clinical dose of the cytotoxic drug alone would usually be appropriate. Further adjustments in the dose of the cytotoxic drug, either up or down, can then be derived from the initial clinical experience.

Radiation and chemotherapy sensitizers

Additional preclinical studies are usually important for the development of sensitizing agents for oncologic indications. In addition to the standard profile of toxicology studies in two species for the sensitizer alone, data on the ability of a sensitizing agent to enhance the toxicity of a cytotoxic or cytostatic therapy to non-neoplastic tissue is highly desirable. As for MDR-reversing agents, a study of the sensitizer combined with the cytotoxic therapy in one species (usually a rodent) that assesses toxicity at both minimally and significantly toxic doses of the cytotoxic agent or radiation therapy is considered important (Table 3). Although this is a straightforward toxicology study when the sensitizer is combined with a drug (e.g. L-buthionine-*S,R*-sulfoximine and alkylating agents), it is not so simple with radiosensitizers because radiation toxicity may only be apparent upon histopathologic examination and because the toxicity can be substantially delayed. One approach to address this issue with radiosensitizers is to conduct skin and leg contracture assays in mice [43] in lieu of comprehensive toxicology studies of the combination. The dosing scheme for these animals should be designed to support the planned clinical trial, but given the common clinical use of highly fractionated radiotherapy, this may not always be feasible.

How best to conduct the initial clinical trial is highly dependent on the combination modality, and advice on dose escalation and scheduling is product-specific. When the sensitizer is intended for combination with a therapy that has curative potential, the starting dose, frequency of dosing, and dose escalation plan for the new sensitizer needs to be carefully considered. Enhanced toxicities from the combination that significantly shorten or delay cycles of the standard therapy should be avoided so that efficacy of the standard therapy is maintained. One

accepted approach is to administer a full standard dose of radiation or anticancer agent, and a dose of the sensitizer projected to have some activity but that imparts little toxicity to the treatment regimen. Other approaches may also be acceptable provided that they are supported by a sound scientific rationale.

Chemoprotection

Chemoprotection is the use of drugs to mitigate the toxic effects of antineoplastic compounds. Marketed examples of this class include dexrazoxane, amifostine, mesna and leucovorin, which decrease the toxicities of doxorubicin (heart), cisplatin (kidney), ifosfamide (bladder) and methotrexate (high dose rescue), respectively. The toxicologic testing of the chemoprotective agent alone in one rodent and one nonrodent species should be based on the proposed use in the clinical trials. Usually these studies are done with a similar route, schedule and duration of administration as when combined with the antineoplastic agent. Reproductive toxicity testing for the protectant alone should be considered when the protectant is to be combined with a chemotherapeutic agent not known to be teratogenic. When the chemotherapeutic agent is known to be teratogenic, it may be useful to assess the ability of the protective agent to prevent this toxicity. The initial clinical dose for a chemoprotectant should ideally be chosen based on projected efficacy, but should not exceed the dose selected by standard toxicity criteria (i.e. one-tenth the rodent STD_{10} unless that dose is severely toxic to nonrodents).

While the primary issue is the toxicity of the chemoprotective agent alone, additional concerns include the possibility of protection of the tumor from the antineoplastic effects of chemotherapy and the possible augmentation of some of the toxic effects of the chemotherapeutic agent. For example, leucovorin, while able to mitigate the effects of an overdose of methotrexate, can also increase the toxicity of 5-fluorouracil [42]. Diethyldithiocarbamate, investigated to decrease the toxicity of cisplatin, actually increased the rate of tumor regrowth following the end of chemotherapy in a rat in vivo model when administered intraperitoneally [3]. In a clinical study, no change in response to 100 mg/m^2 cisplatin was noted, while the patient withdrawal due to toxicity was increased significantly in the diethyldithiocarbamate plus cisplatin arm [24]. In another clinical study, treatment with pyridoxine (vitamin B_6) to reduce the neurotoxicity of hexamethylmelamine and cisplatin was associated with a significant decrease in duration of response in ovarian cancer patients [48]. These clinical findings emphasize that the potential for tumor protective effects and changes in toxicity should be examined in preclinical studies of chemoprotective agents.

Toxicity data on the combination of the chemoprotectant and the antineoplastic agent can be derived from efficacy experiments, provided that histopathologic data are collected. Although in vitro data are useful, the influence on tumor protection by the chemoprotectant should be examined in vivo, where additional factors such as changes in metabolic profile of either drug may affect outcome. Comparisons of the duration of response (i.e. time to tumor regrowth) between an antineoplastic alone and the combination of antineoplastic and chemoprotectant are particularly important. This information on the interaction between the chemoprotective agent and the chemotherapeutic agent can be valuable for the design of pivotal or large scale clinical studies.

Considerations for changes in route or formulation

Changes in the route of administration or in the formulation of anticancer drugs are often pursued with a goal of improving drug utility. If a clinical trial is proposed by the oral route for a drug that has already been investigated by intravenous administration, then additional preclinical studies should address whether there is enhanced liver toxicity, direct gastrointestinal toxicity, or altered metabolism (due to microflora in the gastrointestinal tract, the intestinal wall, or a first-pass effect through the liver). An oral animal toxicity study with bioavailability data or an oral animal efficacy study with assessment of gastrointestinal and liver toxicity can address these concerns. The schedule of administration used in such a study should reflect the planned schedule of administration in the proposed phase I clinical study. A careful assessment of the forms of the drug present in blood should also be attempted, particularly if it is believed that metabolites contribute to the activity or toxicity. Pharmacokinetic information in humans for the i.v. formulation would also be useful for determining the starting dose of the oral formulation, but is not mandatory.

When i.v. administration is proposed for a drug with which there is oral clinical experience, the main concern is that the systemic exposure and resulting toxicity may be much greater by the i.v. route. Either an i.v. animal toxicity study (using the same schedule of administration as proposed for the initial phase I trial) or pharmacokinetic data with the oral formulation in humans that supports an acceptable exposure after the i.v. administration is important before beginning a trial with an i.v. formulation. Similarly, concerns about increased systemic exposure should be addressed when the formulation of an oral anticancer agent in clinical trials is changed. The studies needed to support use of the new formulation depend on the bioavailability of the original formulation in humans and on whether the potential exists to significantly increase bioavailability with the new formulation. For example, if the bioavailability in humans of the original formulation is near 100%, then there is little risk of increased toxicity with the new formulation and no new studies would be needed. On the other hand, if the bioavailability of the original formulation in humans is low, then a bioavailability

study comparing the new formulations in an appropriate species should be considered. An appropriate starting dose for the initial phase I trial with the new formulation can be projected from these data. In some circumstances, it may be sufficient to test a dose of a new formulation in humans without an animal study, so long as the dose is reduced to take into account potential changes in bioavailability.

Summary

Preclinical studies are an essential component of the drug development process. The preclinical development of new anticancer drugs is unique because of the life-threatening nature of the disease and because in most cases humans will be dosed to toxicity. In oncology, these studies are particularly useful in determining potentially safe and effective starting doses and schedules for a clinical trial. These studies also help to predict clinical toxicities and their reversibility, and provide a means for the determination of a dose-escalation scheme. The availability of adequate preclinical data can minimize the number of patients treated with ineffective doses or therapies in phase I trials and allow rapid determination of phase II doses. Preclinical studies are most useful when conducted using the same schedule, duration, formulation, and route of administration as that proposed in the clinical trial.

Basic research continues to provide information about new cellular mechanisms central to malignancy and often leads to drugs that attempt to exploit those mechanisms. The optimal development of a new class of drugs may differ from successful approaches used in the development of older well-established classes. New biological endpoints and new methods in toxicology may also be discovered and cannot be anticipated. The recommendations in this report have thus attempted to avoid being so restrictive and specific as to impede the development of innovative therapeutics for clinical use. Instead, the concerns that should be addressed have been emphasized.

It is assumed that most of the studies conducted to assess the toxicity profile of a drug follow GLP [16, 17]. Before a phase I clinical trial is initiated in patients with advanced cancers, two preclinical toxicity studies are usually conducted. One is a study in a rodent species that can identify doses that result in life-threatening and non-life-threatening toxicities. The other is a study to confirm that doses are identified that are not lethal and do not cause serious or irreversible toxicity in a nonrodent species. These studies, to the extent feasible, should be based on a rational schedule for efficacy and mimic the schedule and duration proposed in the phase I clinical trial.

Although not required, pharmacodynamic and pharmacokinetic studies can provide substantial additional support for the safety profile (starting dose, escalation, and drug combinations) and optimal potential use of the drug (tumor type, schedule, and route). This information is especially important in the development of noncytotoxic drugs (e.g. MDR reversers and immunomodulators) where the objective of the phase I or II clinical study may not be to reach MTD.

Depending on the type of antineoplastic agent under study, different approaches for estimating starting doses are appropriate. The phase I starting dose for cytotoxic agents in humans is generally one-tenth of the rodent STD_{10} on a milligrams per meter squared basis so long as this starting dose does not cause serious irreversible toxicities in nonrodents. If this dose causes irreversible toxicities in nonrodents, then the starting dose should be no more than one-sixth of the highest dose that does not produce lethality or serious irreversible toxicity in the nonrodent species. For noncytotoxic agents, starting dose selection should take into account the drug's pharmacodynamically active doses, provided that they do not cause substantial toxicity. Regardless of the method used to select the starting dose, the planned dose escalation scheme should be designed based on the slopes of the dose response curves for toxicodynamics and pharmacodynamics, the types of toxicities observed, and the pharmacokinetics of the drug.

In the later stages of anticancer drug development, when information is available on toxicity to humans, the need for additional toxicology studies should be evaluated. For most cytotoxic drugs, toxicity studies of limited duration suffice. With drugs intended for chemoprevention, adjuvant therapy, long-term hormone therapy, or long-term immunomodulator therapy, animal toxicity studies up to a maximum of 6 months in rodents and 12 months in a nonrodent species may be important for assessing safety and for supporting marketing approval. In addition, reproductive toxicity and carcinogenicity studies should be conducted when appropriate (e.g. for chemoprevention indications). Depending upon the nature of toxicity profiles in animal species and humans, special studies addressing potential organ system toxicities may also be useful.

The CDER Division of Oncology Drug Products of the FDA welcomes discussion of specific anticancer drugs at the early stages of development to facilitate rapid and efficient drug development. Prior to filing an IND, sponsors may have discussions with appropriate FDA staff and request pre-IND evaluations of their study plan. This may help sponsors to avoid spending time and resources on unnecessary studies, and may help to expedite initiation of clinical studies of promising new drugs.

Acknowledgements We thank Drs. Young Soo Choi, Misoon Chun, Tao Du, Lawrence F. Sancilio, and Virgil Whitehurst of the Division of Pulmonary Drug Products, FDA; Drs. Robert J. DeLap, Robert Justice, and the Medical Officers of the Division of Oncology Drug Products, FDA; and Dr. Robert J. Temple of the Office of Drug Evaluation I, FDA for their helpful comments during the preparation of this report.

References

1. Barbui T, Rambaldi A, Parenzan L, Zucchelli M, Perico N, Remuzzi G (1992) Neurological symptoms and coma associated with doxorubicin administration during chronic cyclosporin therapy. Lancet 339: 1421
2. Bartlett NL, Lum BL, Fisher GA, Brophy NA, Ehsan MN, Halsey J, Sikic BI (1994) Phase I trial of doxorubicin with cyclosporine as a modulator of multidrug resistance. J Clin Oncol 12: 835
3. Borch RF, Katz JC, Lieder PH, Pleasants ME (1980) Effect of diethyldithiocarbamate rescue on tumor response to cisplatinum in a rat model. Proc Natl Acad Sci USA 77: 5441
4. Branda RF (1981) Transport of 5-methyltetrahydrofolic acid in erythrocytes from various mammalian species. J Nutr 111: 618
5. Branham WS, Zehr DR, Chen JJ, Sheehan DM (1988) Alterations in developing rat uterine cell populations after neonatal exposure to estrogens and antiestrogens. Teratology 38: 271
6. Choudary J, Contrera JF, DeFelice A, DeGeorge JJ, Farrelly JG, Fitzgerald G, Goheer MA, Jacobs A, Jordan A, Meyers L, Osterberg R, Resnick C, Sun CJ, Temple RJ (1996) Response to Monro and Mehta proposal for use of single-dose toxicology studies to support single-dose studies of new drugs in humans. Clin Pharmacol Ther 59: 265
7. Collins JM, Zaharko DS, Dedrick RL, Chabner BA (1986) Potential roles for preclinical pharmacology in phase I clinical trials. Cancer Treat Rep 70: 73
8. Collins JM, Grieshaber CK, Chabner BA (1990) Pharmacologically guided phase I clinical trials based upon preclinical drug development. J Natl Cancer Inst 82: 1321
9. Cordon-Cardo C, O'Brien JP, Boccia J, Casals D, Bertino JR, Melamed MR (1990) Expression of the multidrug resistance gene product (P-glycoprotein) in human normal and tumor tissues. J Histochem Cytochem 38: 1277
10. DeVita VT Jr (1993) Principles of chemotherapy. In: DeVita VT Jr, Hellman S, Rosenberg SA (eds) Cancer: principles and practice. Lippincott, Philadelphia, p 276
11. Doroshow JH, Locker GY, Myers CE (1979) Experimental animal models of adriamycin cardiotoxicity. Cancer Treat Rep 63: 855
12. Dougherty TJ (1993) Photodynamic therapy. Photochem Photobiol 58: 895
13. Faries D (1994) Practical modifications of the continual reassessment method for phase I cancer clinical trials. J Biopharm Stat 4: 147
14. Food and Drug Administration (1994) International Conference on Harmonization. Guideline for detection of toxicity to reproduction for medicinal products. Fed Register 59: 48746
15. Food and Drug Administration (1995) International Conference on Harmonization. Toxicokinetics: the assessment of systemic exposure in toxicity studies. Fed Register 60: 11264
16. Food and Drug Administration (1995) Content and format of investigational new drug applications (INDs) for phase 1 studies of drugs, including well-characterized, therapeutic, biotechnology-derived products. Center for Drug Evaluation and Research, Rockville, Md.
17. Food and Drug Administration (1996) Good laboratory practice for nonclinical laboratory studies. 21 Code Federal Regulations 58: 265
18. Food and Drug Administration (1996) International Conference on Harmonization. Final guideline on the need for long-term rodent carcinogenicity studies of pharmaceuticals. Fed Register 61: 8154
19. Food and Drug Administration (1996) International Conference on Harmonization. Guidance on specific aspects of regulatory genotoxicity tests for pharmaceuticals. Fed Register 61: 18198
20. Food and Drug Administration (1996) International Conference on Harmonization. Draft guideline on testing for carcinogenicity of pharmaceuticals. Fed Register 61: 43297
21. Food and Drug Administration (1997) Points to consider in the manufacture and testing of monoclonal antibody products for human use. Center for Biologics Evaluation and Research
22. Food and Drug Administration (1997) International Conference on Harmonization. Genotoxicity: a standard battery for genotoxicity testing of pharmaceuticals. Fed Register 62: 16025
23. Freireich EJ, Gehan EA, Rall DP, Schmidt LH, Skipper HE (1966) Quantitative comparison of toxicity of anticancer agents in mouse, rat, hamster, dog, monkey, and man. Cancer Chemother Rep 50: 219
24. Gandara DR, Nahhas WA, Adelson MD, Lichtman SM, Podczaski ES, Yanovich S, Homesley HD, Braly P, Ritch PS, Weisberg SR, Williams L, Diasio RB, Perez EA, Karp D, Reich SD, McCarroll K, Hoff JV (1995) Randomized placebo-controlled multicenter evaluation of diethyldithiocarbamate for chemoprotection against cisplatin-induced toxicities. J Clin Oncol 13: 490
25. Giusti RM, Iwamoto K, Hatch EE (1995) Diethylstilbestrol revisited: a review of the long-term health effects. Ann Intern Med 122: 778
26. Gottesman MM, Pastan I (1993) Biochemistry of multidrug resistance mediated by the multidrug transporter. Annu Rev Biochem 62: 385
27. Graham MA, Workman P (1992) The impact of pharmacokinetically guided dose escalation strategies in phase I clinical trials: critical evaluation and recommendations for future studies. Ann Oncol 3: 339
28. Gralla EJ, Fleischman RW, Luthra YK, Stadnicki SW (1979) The dosing schedule dependent toxicities of adriamycin in beagle dogs and rhesus monkeys. Toxicology 13: 263
29. Grieshaber CK, Marsoni S (1986) Relation of preclinical toxicology to findings in early clinical trials. Cancer Treat Rep 70: 65
30. Hard GC, Iatropoulos MJ, Jordan K, Radi L, Kaltenberg OP, Imondi AR, Williams GM (1993) Major difference in the hepatocarcinogenicity and DNA adduct forming ability between toremifene and tamoxifen in female Crl:CD(BR) rats. Cancer Res 53: 4534
31. Iguchi T, Hirokawa M, Takasugi N (1986) Occurrence of genital tract abnormalities and bladder hernia in female mice exposed neonatally to tamoxifen. Toxicology 42: 1
32. Kelloff GJ, Johnson JR, Crowell JA, Boone CW, DeGeorge JJ, Steele VE, Mehta MU, Temeck JW, Schmidt WJ, Burke G, Greenwald P, Temple RJ (1995) Approaches to the development and marketing approval of drugs that prevent cancer. Cancer Epidemiol Biomarkers Prev 4: 1
33. Kim S, Kim DJ, Howell SB (1987) Modulation of the peritoneal clearance of liposomal cytosine arabinoside by blank liposomes. Cancer Chemother Pharmacol 19: 307
34. Lelieveld P, Vijgh WJF van der, Velzen D van (1987) Preclinical toxicology of platinum analogues in dogs. Eur J Cancer Clin Oncol 23: 1147
35. Lowe MC, Davis RD (1984) The current toxicology protocol of the National Cancer Institute. In: Helmann K, Carter SK (eds) Fundamentals of cancer chemotherapy. McGraw-Hill, New York, p 228
36. Mettler FP, Young DM, Ward JM (1977) Adriamycin-induced cardiotoxicity (cardiomyopathy and congestive heart failure) in rats. Cancer Res 37: 2705
37. Montandon F, Williams GM (1994) Comparison of DNA reactivity of the polyphenylethylene hormonal agents diethylstilbestrol, tamoxifen and toremifene in rat and hamster liver. Arch Toxicol 68: 272
38. National Cancer Institute (1993) Investigator's handbook, Bethesda, Md.
39. O'Quigley J, Pepe M, Fisher L (1990) Continual reassessment method: a practical design for phase I clinical trials in cancer. Biometrics 46: 33
40. Peck CC, Barr WH, Benet LZ, Collins J, Desjardins RE, Furst DE, Harter JG, Levy G, Ludden T, Rodman JH, Sanathanan L, Schentag JJ, Shah VP, Sheiner LB, Skelly JP, Stanski DR, Temple RJ, Viswanathan CT, Weissinger J, Yacobi A (1992) Opportunities for integration of pharmacokinetics, pharmacodynamics, and toxicokinetics in rational drug development. Clin Pharmacol Ther 51: 465
41. Philip PA, Rea D, Thavasu P, Carmichael J, Stuart NSA, Rockett H, Talbot DC, Ganesan T, Pettit GR, Balkwill F,

Harris AL (1993) Phase I study of bryostatin 1: assessment of interleukin 6 and tumor necrosis factor alpha induction in vivo. J Natl Cancer Inst 85: 1812
42. Poon MA, O'Connell MJ, Wieand HS, Krook JE, Gerstner JB, Tschetter LK, Levitt R, Kardinal CG, Maillard JA (1991) Biochemical modulation of fluorouracil with leucovorin: confirmatory evidence of improved therapeutic efficacy in advanced colorectal cancer. J Clin Oncol 9: 1967
43. Stone HB (1984) Leg contracture in mice: an assay of normal tissue response. Int J Radiat Oncol Biol Phys 10: 1053
44. Taguchi O (1987) Reproductive tract lesions in male mice treated neonatally with tamoxifen. Biol Reprod 37: 113
45. Taguchi O, Nishizuka Y (1985) Reproductive tract abnormalities in female mice treated neonatally with tamoxifen. Am J Obstet Gynecol 151: 675

Note added in proof The draft ICH S4 document "Duration of chronic toxicity testing in animals (rodent and nonrodent toxicity testing)" is under consideration by the EU, Japan, and US. If implemented in its current form, the maximum duration of toxicity testing for nonrodents would change from 12 months to 9 months for most drugs undergoing international development.

46. Taudy M, Syka J, Popelar J, Ulehlova L (1992) Carboplatin and cisplatin ototoxicity in guinea pigs. Audiology 31: 293
47. US Interagency Staff Group on Carcinogens (1986) Chemical carcinogens: a review of the science and its associated principles. Environ Health Perspect 67: 201
48. Wiernik PH, Yeap B, Vogl SE, Kaplan BH, Comis RL, Falkson G, Davis TE, Fazzini E, Cheuvart B, Horton J (1992) Hexamethylmelamine and low or moderate dose cisplatin with or without pyridoxine for treatment of advanced ovarian carcinoma: a study of the Eastern Cooperative Oncology Group. Cancer Invest 10: 1
49. Witt PL, Ritch PS, Reding D, McAuliffe TL, Westrick L, Grossberg SE, Borden EC (1993) Phase I trial of an oral immunomodulator and interferon inducer in cancer patients. Cancer Res 53: 5176

1

How Oncology Drug Development Differs from Other Fields

Ann T. Farrell, John Leighton, Grant Williams, and Richard Pazdur

United States Food and Drug Administration, Division of Oncology Drug Products, Center for Drug Evaluation and Research, Rockville, Maryland 20852

DRUG DEVELOPMENT

Drug development is a stepwise process progressing from preclinical to clinical evaluation. Drug development programs differ because the "risk-benefit" ratio for a drug depends on the target disease and the patient population. Cancer is an aggressive and potentially fatal disease for which the patient receives therapy conventionally administered at or near maximally tolerated doses. The life-threatening nature of advanced cancer allows the acceptance of considerably more risk than would be acceptable in other conditions. Even within oncology drug development, differences exist for drugs being developed for chemoprevention compared with drugs being developed for the treatment of refractory disease because both the potential benefit of therapy and the acceptability of risk vary in different cancer settings. Oncology chemopreventive drug development more closely resembles that of drug development for non–life-threatening disease. Requirements for the preclinical and clinical data for oncology drugs also differ from those for other drugs.

General Considerations for Preclinical Drug Development

The goal of the preclinical safety evaluation is to ensure adequate characterization of toxic effects with respect to target organs, potential reversibility, dose dependence, and relationship to exposure (1). These preclinical studies should be conducted under Good Laboratory Practices. Preclinical studies determine the choice of an initial starting dose for clinical studies (2). The preclinical safety evaluation also identifies the potential organ toxicities to be monitored in the clinical studies. Serious adverse events observed during preclinical or clinical studies may warrant additional specific safety studies.

Toxicity Studies

Single-dose (acute) toxicity is usually assessed in two mammalian species (generally one rodent and one nonrodent species) (1,3). A dose-escalation study with an appropriate toxicity evaluation may substitute for an acute single-dose toxicity study (1). Toxicity studies that use doses high enough to cause toxicity assess clinical signs, body weight, food consumption, gross pathology, and histopathology. Repeat dose toxicity studies should use regimens similar to those planned for the clinical studies, including schedule, duration, and route of administration (3,4). If the drug will be administered subcutaneously, intramuscularly, dermally, or ophthalmologically, preclinical local tolerance studies should be performed, including single and repeat dose, if necessary (5).

National drug regulatory agencies differ somewhat concerning the preclinical data that are recommended prior to human use. The next two tables outline the duration of repeat-dose toxicity studies expected as support for clinical trials in the United States, the European Union, and Japan (1,3,6–9) (Tables 1.1 and 1.2).

Toxicokinetic Studies

Toxicokinetic studies generate pharmacokinetic data, which describe the systemic exposure achieved in

The views expressed are the result of independent work and do not represent the views or findings of the United States Food and Drug Administration or the United States Government.

TABLE 1.1. *Duration of repeated-dose toxicity studies to support phase 1 and 2 trials in the European Union and phase 1, 2, and 3 trials in the United States and Japan*

	Minimum duration of repeated-dose toxicity studies	
Duration of clinical trials[a]	Rodents	Nonrodents
Single dose	2–4 wk	2 wk
Up to 2 wk	2–4 wk	2 wk
Up to 1 mo	1 mo	1 mo
Up to 3 mo	3 mo	3 mo
Up to 6 mo	6 mo	6 mo
> 6 mo	6 mo	Long-term

In Japan, if there are no phase 2 clinical trials of equivalent duration to the planned phase 3 trials, conduct of longer-duration toxicity studies should be considered as presented in Table 1.2. In the European Union and the United States, 2-week studies are the minimum duration. In Japan, 2-week nonrodent and 4-week rodent studies are needed. In the United States, as an alternative to 2-week studies, single-dose toxicity studies with extended examinations can support single-dose human trials.

[a] Data from 6 months of administration in nonrodents should be available before the initiation of clinical trials longer than 3 months. Alternatively, if applicable, data from a 9-month nonrodent study should be available before the treatment duration exceeds that which is supported by the available toxicity studies.

Revised from the International Conference on Harmonization Document M3. Guidance for industry: nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. Geneva: ICH, July 1997.

animals and its relationship to dose level, to any observed toxicity and the time course of the development of toxicity. This information may subsequently be used to adjust the choice of species for further study and for planning dosing schedule and study design in clinical trials. Pharmacokinetic data may be derived either from specially designed studies or from planned nonclinical toxicity studies, such as single-dose, repeat-dose, *in vivo* genotoxicity, carcinogenicity, or reproductive toxicity studies (10). Reproductive preclinical toxicokinetic data collection may involve exposure assessment in dams, embryos, fetuses, or the newborn. Tissue distribution studies provide information on the distribution and accumulation of the drug and its metabolites. Repeated-dose tissue distribution studies may be required under certain circumstances (11). Toxicokinetic study data should be available when clinical phase 1 studies are completed (1).

TABLE 1.2. *Duration of repeated-dose toxicity studies to support phase 3 trials in the European Union and marketing in Japan, the European Union, and the United States*

	Minimum duration of repeated-dose toxicity studies	
Duration of clinical trials	Rodents	Nonrodents
Up to 2 wk	1 mo	1 mo
Up to 1 mo	3 mo	3 mo
Up to 3 mo	6 mo	3 mo
> 3 mo	6 mo	Long-term

These data also reflect the marketing recommendations in Japan, the European Union, and the United States except that a long-term nonrodent study is recommended for clinical use exceeding 1 month.

Revised from the International Conference on Harmonization Document M3. Guidance for industry: nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. Geneva: ICH, July 1997.

Genotoxicity

Genotoxicity testing has been used to predict carcinogenicity. Pharmaceutical agents testing positive in genotoxicity tests have the potential to be human carcinogens and/or mutagens, which may induce cancer. In general, *in vitro* and *in vivo* genotoxicity testing detects direct or indirect genetic damage and should be performed prior to the initiation of phase 1 studies (6,12). However, in clinical studies of patients with end-stage disease, genetic testing may not be necessary at this stage of drug development. The standard battery of genotoxic tests includes the following: (a) a bacterial reverse mutation assay, (b) an *in vitro* test with cytogenetic evaluation of chromosomal damage with mammalian cells or *in vitro* mouse lymphoma thymidine kinase assay, and (c) an *in vivo* test for chromosomal damage using rodent hematopoietic cells. Negative results obtained with compounds to all three types of studies suggest the absence of genotoxicity. The standard battery may be modified when drugs are excessively toxic to bacteria, resulting in interference with the bacterial reverse mutation assay or with mammalian cell replication. In these situations, the performance of two mammalian cell *in vitro* tests using different cell types and different endpoints (e.g., chromosomal damage and gene mutation) should be considered.

Compared with *in vitro* testing, *in vivo* testing provides information on the absorption, distribution, metabolism, and excretion of the drug. The utility of *in vivo* testing is limited with agents that are not sufficiently absorbed, such as radioimaging agents. A single positive result in a genotoxicity assay does not necessarily imply that a genotoxic risk exists (12). Genotoxic test interpretation may be complicated by the concentration, culture conditions, and reproducibility of test results in drugs of the same or similar class. If *in vitro* genotoxicity results are equivocal or positive, additional *in vitro* or *in vivo* testing may be required.

Carcinogenicity Studies

Carcinogenicity studies identify tumorigenic potential in animals and evaluate possible human risk. These studies are performed primarily when the drug is expected to be administered regularly over a substantial part of a patient's life. Typically, these studies dose animals using the same administration route intended for humans. Several doses are tested. In the past, differences among international regulatory authorities existed concerning high dose selection. All regulatory authorities currently use a maximum feasible dose.

For most drugs used to treat advanced cancer, carcinogenicity studies may be completed later in drug development, including in the postapproval setting. In disease populations where the life expectancy is short, such as metastatic cancer, long-term carcinogenicity studies may not be required. However, in clinical situations where therapies are curative or have a pronounced effect on survival (e.g., patients receiving adjuvant chemotherapy or hormonal therapy to prevent cancer recurrence), carcinogenicity studies are usually necessary. Similarly, carcinogenicity studies are necessary for chemopreventive agents.

Carcinogenicity studies may be required if (a) the anticipated use of the drug is 3 to 6 months or longer, (b) concern exists about carcinogenic potential (e.g., carcinogenicity test results have been positive for other drugs in the product class, (c) the intended patient population has a life expectancy greater than 2 or 3 years, (d) the agent is an ophthalmologically or dermally applied product that may have extensive systemic exposure, and (e) the drug product is similar to an endogenous substance given as replacement therapy.

Completed carcinogenicity studies are usually not required when the pharmaceutical agent is unequivocally genotoxic (13). Regulatory authorities have usually required two long-term carcinogenicity rodent studies prior to marketing of a new drug (13). Regulatory authorities have recently considered the utility of one long-term and one short- or medium-term *in vivo* rodent test systems designed to clarify a particular carcinogenicity concern (e.g., initiation-promotion models in rodents) (14). Several regulatory guidances provide the study design, necessary monitoring, and required investigations (14–17). Mechanistic studies (e.g., specialized genotoxicity studies) may be useful in the interpretation of tumor findings in carcinogenicity studies (14).

Reproductive Toxicity Studies

Preclinical reproductive toxicity studies are used to investigate the drug's effect on mammalian reproduction (18). The necessity for these studies is based on the drug's anticipated use in relation to reproductive life cycle. Rats are the predominant species used. Embryotoxicity studies, however, require two species testing. Usually rabbits are the second mammalian species used for embryotoxicity studies because of the extensive prior background knowledge of this species. Reproductive toxicity testing is performed with study designs that focus on three periods: (a) fertility to early embryonic development (premating to conception to implantation), (b) prenatal to postnatal

development, including maternal function (implantation to birth to sexual maturity), and (c) embryo/fetal development (implantation to end of pregnancy).

In the United States and European Union, men may be included in phase 1 and 2 clinical studies prior to a preclinical assessment of male fertility if an assessment of male reproductive organs was performed in repeat-dose toxicity studies (1,19). In Japan, prior to inclusion of men in studies, preclinical male fertility studies are usually performed. Ideally, preclinical male fertility testing should be performed prior to initiation of phase 3 trials.

Women who are not of childbearing potential can be included in clinical trials prior to completion of reproductive toxicity provided that an assessment of the female reproductive tract was performed in repeat-dose toxicity studies. In the United States, women of childbearing potential can be included in early studies without reproductive toxicity studies provided that highly effective birth control or pregnancy testing is used to minimize the risk (1). Highly effective birth control is defined as a method that results in a low failure rate (i.e., less than 1%) when appropriately used. Continued pregnancy testing and birth control compliance monitoring during the trial is performed to minimize risk. Informed consent forms should discuss potential risk, especially if no information is known.

In Japan, prior to the inclusion of women of childbearing potential using contraception in any trial, female fertility and embryo/fetal development testing should be performed. In the European Union, embryo/fetal assessment should be performed and completed prior to the inclusion of women of childbearing potential in phase 1 studies. In the United States, preclinical assessment of female fertility and embryo/fetal development should be performed prior to entering women of childbearing potential into phase 3 trials, even if the women are using adequate birth control. In the European Union, female fertility studies should be performed prior to phase 3. Preclinical prenatal and postnatal data should be available prior to marketing approval. For all three drug development regions, all female reproductive toxicity and standard battery of genotoxicity studies should be completed prior to the inclusion of women of childbearing potential who are not using highly effective birth control methods, or whose pregnancy status is uncertain.

Safety Pharmacology Studies

Safety pharmacology studies investigate the potential undesirable pharmacodynamic effects of a drug on physiologic function (20,21). These studies have three goals: (a) to identify undesirable pharmacodynamic effects of a drug on physiologic function, possibly relating to safety, (b) to evaluate adverse effects (e.g., pathophysiologic and/or pharmacodynamic) observed in toxicology and/or clinical studies, and (c) to investigate the mechanism of the observed or suspected adverse events. The most important adverse events to investigate are those that effect critical and essential functions (central nervous system, cardiovascular, and respiratory systems). Adverse events transiently effecting the hepatic, renal, or gastrointestinal system that do not cause irreversible effect may not warrant immediate study, except when there may be irreparable harm in a specific vulnerable patient population (e.g., gastrointestinal toxicity in an agent being developed for Crohn's disease). These studies may involve *ex vivo* and *in vitro* testing with isolated organs, tissues, cell cultures, cellular fragments, receptors, ion transporters, and enzymes (22).

Oncologic versus Nononcologic Preclinical Drug Development

Due to the life-threatening nature of an unresectable malignancy, conventional therapeutic strategies employ anticancer drugs, radiation therapy, and surgery with significant toxicity and complications, different risk/benefit considerations are used in regulatory decision making for cancer therapies (2,23). Preclinical studies for drugs to treat serious and life-threatening conditions may be abbreviated, deferred, or even omitted (1). Oncology preclinical drug development is tailored according to the intended use of the drug and the eventual patient population in the indication.

Starting Dose Determination for Traditional Cytotoxic Oncology Drugs

Like nononcology drugs, investigational cancer drugs should have two toxicology studies to support initial phase 1 clinical studies. One study is a required rodent study, which identifies the doses producing life-threatening and non–life-threatening toxicities. The rodent study identifies the dose (mg/m^2) that is severely toxic to 10% of the rodents (STD_{10}). If no life-threatening toxicity is observed in the rodent study, then the highest dose tested is taken to be the STD_{10}. The second study should confirm that this dose (STD_{10}) does not cause irreversible toxicity in nonrodents. The starting dose for a phase 1 study is one tenth of the STD_{10}. If life-threatening toxicity is observed in the nonrodent study at the proposed starting dose based on rodent studies, then the starting dose (mg/m^2) for clinical studies is one sixth of the

highest dose tested in nonrodents that does not produce severe toxicity.

Dose Determination for Noncytotoxic Oncology Drugs

Many noncytotoxic oncology drugs use the approach for cytotoxic drugs to determine the starting dose. Preclinical pharmacodynamic information can facilitate dose determination in oncologic phase 1 drug development. The determination of the pharmacodynamically active dose may be more useful than maximum tolerated dose (MTD) determination for immunomodulators and noncytotoxic agents. For example, no additional benefit may occur from the administration of doses higher than that necessary to successfully block a receptor. Additional higher drug doses beyond the pharmacologically identified dose may incur toxicity without additional patient benefit.

Preclinical Data Required Prior to Investigational New Drug Application Filing

Preclinical studies necessary at the time of an Investigational New Drug (IND) and a New Drug Application (NDA) filings vary depending upon the drug product, proposed indication and patient population, the observed clinical outcome, characterization of toxicities observed in animals and humans, and the projected treatment duration. If a nononcologic drug is developed for a serious or life-threatening condition, the development may be similar to an oncologic drug. At the initial IND filing, the following preclinical studies are recommended: single-dose toxicity, repeat-dose toxicity, and genotoxicity. Genotoxicity studies may not be required at the IND filing for studies conducted in patients with advanced cancer. Additional preclinical studies may be required prior to IND filing for photosensitizing agents, antibody conjugate, liposomal delivery, and cytotoxics delivered via depot (2).

Preclinical Considerations for Cytotoxics

For cytotoxics prescribed in advanced disease, the necessary preclinical repeat-dose toxicity study may be only 28 days in duration. For cytotoxics used in the adjuvant setting, multiple-cycle toxicity studies may be required, especially if clinical data regarding repeat administration in a more advanced cancer/metastatic disease population do not exist. Preferentially, data from International Conference on Harmonization (ICH) stage C-D teratogenicity studies should be available prior to NDA submission but are not required prior to submission of a phase 1 protocol for treatment of advanced cancer. Carcinogenicity studies are usually not required for cytotoxics to treat advanced cancer, but may be required for these drugs used in adjuvant treatment, especially if the treatment duration would be longer than 6 months. Special safety pharmacology studies may be required to support an NDA filing (e.g., special cardiac studies for anthracyclinelike agents).

The initial evaluation of cytotoxic combination therapy has traditionally been performed in the clinical setting. This empirical approach has been relatively successful but may not be optimal. Preclinical testing of combinations provides the opportunity to explore various doses, dose ratios, schedules, and drug sequencing. Preclinical combination testing may not be necessary provided that each agent's toxicities have been fully characterized, and data do not indicate that the combination use would be unsafe. Concern may increase if one agent interferes with the metabolism or elimination of another, or if both cytotoxic agents target the same metabolic or cellular pathway, or cellular function. Additional safety testing may be required for the development of photosensitizers or for novel drug-delivery systems (e.g., copolymer implants, human albumin microspheres, monoclonal antibody–drug conjugates, and liposomal encapsulation).

Preclinical Considerations for Noncytotoxic, Chronically Administered Oncologic Drugs

Noncytotoxic, chronically administered oncologic agents (e.g., chemopreventives, hormones, and immunomodulators) should have repeat-dose toxicity studies. The duration will depend on the intended population because the development of these agents is similar to nononcologic drugs. In patients with advanced cancer, 28-day toxicology studies may be sufficient prior to phase 1 or 2 studies. Studies up to 6 months' duration in rodents and up to 12 months in nonrodents may be required prior to clinical studies involving high-risk or cancer-free patients, or those expected to have prolonged survival. The requirement for reproductive toxicity testing depends on the disease stage and intended patient population. ICH stage C-D (developmental) reproductive toxicity studies are important for NDA filing. In addition, ICH stage A-B (premating to implantation) reproductive toxicity studies are requisite for the development of hormonal agents and ICH A-B, C-F (implantation to offspring sexual maturity) reproductive toxicity studies are necessary for the development of chemopreventive agents. Carcinogenicity studies are usually not required for advanced disease, but would be required if the agent were administered to patients who are high-risk, cancer-free, or expected to have prolonged survival.

Preclinical Considerations for Chemotherapy Modulators

The development plan of chemotherapy modulators encompasses determining the modulator's starting dose and toxicities when given alone and the modulator's dose and toxicities when combined with other therapies (e.g., chemotherapy, radiation therapy). Prior to IND submission, single- and repeat-dose toxicity studies should be performed with the modulator alone and combined with intended agents. One study arm should replicate the intended treatment schedule. Genotoxicity testing prior to IND filing is needed in selected occasions when the modulator may be administered to healthy volunteers or patients believed to be cancer-free. If a chemoprotectant is being developed, the toxicology studies should include a histopathologic examination to evaluate the combination's toxicity and should evaluate the possibility of tumor protection. Toxicology studies of sufficient duration for the intended duration in the patient population should be conducted.

Phases of Clinical Drug Development

Phase 1

Phase 1 studies are the initial studies evaluating a new investigational agent in humans (24). For drugs developed for nonserious and non–life-threatening conditions, the phase 1 studies are conducted in normal, healthy volunteers. Phase 1 studies determine the drug's toxicity profile, pharmacodynamic/pharmacokinetic parameters, and the range of nontoxic doses for subsequent trials. In normal subjects, the highest allowed doses are limited by the highest nontoxic doses determined from preclinical animal data.

For drugs developed for serious and life-threatening conditions, the phase 1 studies are usually conducted in patients with the medical condition. The phase 1 studies identify acceptable doses, toxicities, and also provide initial signs of activity. The major goal of phase 1 oncology studies of cytotoxic drugs is to determine the MTD and the dose for subsequent phase 2 testing. A variety of dose-escalating designs exist (e.g., standard or modified Fibonacci, modified continual reassessment, accelerated titration, and pharmacokinetically guided dose strategies) (25,26). The highest doses used in clinical oncology studies are not restricted to those comparative doses used in the preclinical studies as long as the toxicities are easily monitored, reversible, and sufficiently precede lethality in animals. The maximum dose is usually defined by toxicities that are measured by accepted criteria, such as the National Cancer Institute Common Toxicity Criteria.

Prior to allowing a study to proceed, the US Food and Drug Administration (FDA) carefully evaluates the protocols, including starting dose and schedule, data (e.g., preclinical data, prior human use), dose-escalation plan, intended duration of administration, safety-monitoring plan, and the MTD definition. Deviations from common clinical study practices may be allowed provided safety is assured.

Phase 2

Phase 2 studies evaluate the drug's effectiveness for a specific disease and further describe its toxicities. In nononcologic fields, these studies may be dose-ranging, active, or placebo-comparator studies. Phase 2 studies serve as a template for the pivotal phase 3 studies. These nononcologic studies may be designed using the same inclusion and exclusion criteria, efficacy endpoints, and analysis plan as the phase 3 pivotal trials. In oncology, traditional phase 2 studies determine if cytotoxic drugs have activity against a particular tumor type and whether that activity justifies further drug development.

Phase 3

Phase 3 clinical trials are the confirmatory studies in a larger patient population and are intended to gather comparative effectiveness and safety information required to assess the drug's overall risk/benefit relationship. These trials are performed after preliminary evidence of effectiveness has been obtained from phase 2 studies. In most nononcology fields, these studies mimic the design and analysis of the phase 2 studies. In oncology, these phase 3 studies usually provide the first comparison of the drug to the standard treatment. Occasionally, oncology agents may be compared to best supportive care.

IND Exemptions for Nononcology and Oncology Studies

Lawfully marketed drugs or biological products may be exempt from filing an IND if the following criteria are met: (a) the trial is not intended to support FDA approval of a new indication or a significant change in the product labeling, (b) the trial is not intended to support a significant change in the advertising for the product, and (c) the clinical study does not involve a new administration route or dosage level or use in a patient population or other factor that significantly increases the risks (or decreases the acceptability of the

risks) (27). The study must be conducted in accordance with institutional review board and informed consent regulations [21 Code of Federal Regulations (CFR) parts 50 and 56], and conducted in compliance with 21 CFR 312.7 regarding the promotion and charging for investigational drugs. The investigator may determine whether a study is exempt from submission. Although a study may be exempt from IND regulations, it still must be conducted with institutional review board oversight and must comply with the informed consent regulations (21 CFR parts 56 and 50).

Review and Approval Process

The NDA review and approval process ensures that safe and effective drugs are available for the American public. Approval occurs after the Agency has carefully considered the submitted preclinical and clinical study reports, manufacturing data, and a proposed product label. An NDA approval is for one or more specific indications. Once an NDA is approved, the pharmaceutical company may advertise and promote the drug's use consistent with the approved package insert.

A successful NDA must provide substantial evidence of effectiveness derived from adequate and well-controlled studies. The data must demonstrate that the drug is safe for its intended use. The application should define the appropriate patient population and provide adequate information to enable the product's safe and effective use.

Oncology Drug Approval

Recognizing the need for early access to promising medicines for the treatment of cancer and other serious and life-threatening diseases, the Agency has developed policies and procedures to facilitate drug development. These policies and procedures include: Accelerated Approval (28,29), Fast Track Program and Rolling NDA Submission (30), Priority Review (31), and Special Protocol Assessment (32).

Accelerated approval is granted for a new drug that provides benefit over available therapy for diseases that are serious or life threatening (29). Accelerated approval can be granted on a surrogate endpoint "reasonably likely to predict clinical benefit." Regulations mandate that sponsors subsequently demonstrate clinical benefit with due diligence after approval. Response rate is an example of a surrogate endpoint for clinical benefit in the treatment of solid tumors. Full approval could be granted when clinical benefit is demonstrated through an improvement in survival or symptom benefit in subsequent trials. Oncology applications receiving accelerated approval based on a surrogate endpoint in a refractory-disease population may subsequently demonstrate clinical benefit in an earlier stage of disease (e.g., accelerated approval in second-line metastatic breast cancer with full approval in first-line or in the adjuvant setting). Other regulatory authorities such as the European Agency for the Evaluation of Medicinal Products have published guidances on anticancer drug development and approval (33,34).

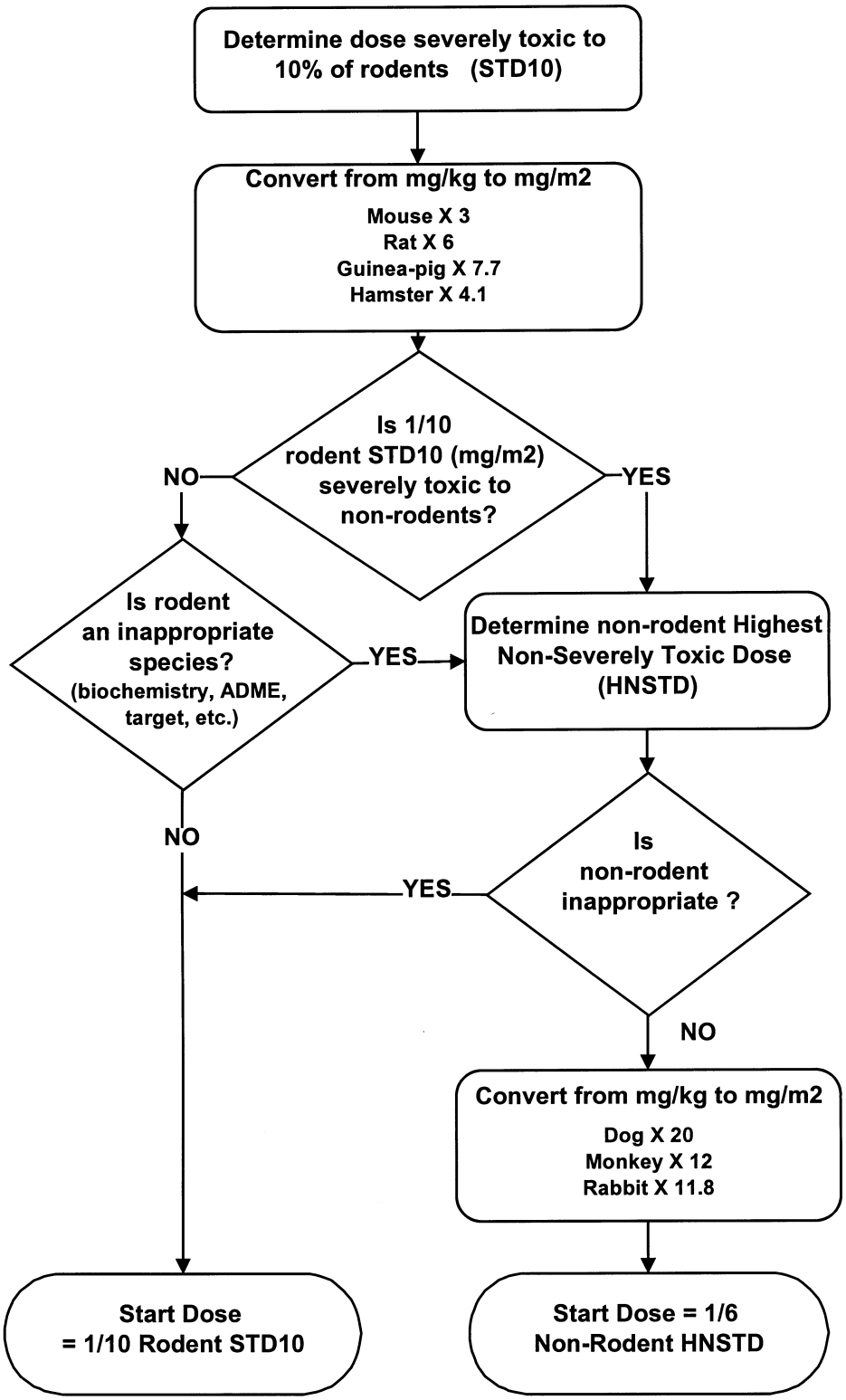
Drug development in oncology differs from other clinical disciplines because of a different risk/benefit relationship. Cancer is a potentially life-threatening disease that requires aggressive measures for treatment and justifies accepting increased toxicity in order to achieved increased efficacy. Drug development in oncology differs from other fields because of the breadth of agents being developed (e.g., cytotoxics, cytostatics, modulators, and chemopreventives). Because of the need to expedite oncology drug development, the quantity and types of preclinical and clinical data may differ from other therapeutic areas. Preclinical and clinical data requirements are tailored to the intended use of the drug and to the population who will eventually use this drug. Differences in preclinical and clinical study requirements may exist among the United States, Japan, and the European Union for the same pharmaceutical product being developed. Continuing consultation with regulatory authorities throughout the drug development process will expedite the approval process for oncology drugs.

REFERENCES

3. International Conference on Harmonization. Guidance for industry: nonclinical safety studies for the conduct of human clinical trials for pharmaceuticals. Geneva: ICH; July 1997. Document M3.
4. [illegible] George JJ, Ahn CH, Andrews P, et al. Regulatory considerations for preclinical development of anticancer drugs. *Cancer Chemother Pharmacol* 1998;41: [illegible]35.
5. [illegible] Department of Health and Human Services, Food and Drug Administration. Single dose acute toxicity testing for pharmaceuticals: revised guidance. *Federal Register* 1996 Aug 26;61:43934–43935.
4. The European Agency for the Evaluation of Medicinal Products Committee for Proprietary Medicinal Products. Note for guidance on repeated dose toxicity. London: EMEA; July 2000.
5. The European Agency for the Evaluation of Medicinal Products Committee for Proprietary Medicinal Products. Note for guidance on non-clinical local tolerance testing of medicinal products. London: EMEA; March 2001.

6. International Conference on Harmonization. Guideline for industry: specific aspects of regulatory genotoxicity tests for pharmaceuticals. Geneva: ICH; April 1996; Document S2A.
7. Choudary J, Contrera JF, DeFelice A, et al. Response to Monro and Mehta proposal for use of single-dose toxicology studies to support single-dose studies of new drugs in humans. *Clin Pharmacol Ther* 1996; 59:265–267.
8. International Conference on Harmonization. Guidance for industry: duration of chronic toxicity testing in animals (rodent and nonrodent). Geneva: ICH; July 1999; Document S4A.
9. US Department of Health and Human Services, Food and Drug Administration. International Conference on Harmonization: guidance on the duration of chronic toxicity testing in animals (rodent and nonrodent toxicity testing) availability. *Federal Register* 1999 Jun 25;64: 43934–43935.
10. International Conference on Harmonization. Guideline for industry toxicokinetics: the assessment of systemic exposure in toxicity studies. Geneva: ICH; March 1995; Document S3A.
11. International Conference on Harmonization. Guideline for industry pharmacokinetics: guidance for repeated tissue distribution. Geneva: ICH; March 1995; Document S3B.
12. International Conference on Harmonization. Guidance for industry: a standard battery for genotoxicity testing of pharmaceuticals. Geneva: ICH; July 1997; Document S2B.
13. International Conference on Harmonization Document S1A Document "Guideline for Industry: The Need for Long-term Rodent Carcinogenicity Studies of Pharmaceuticals." Geneva: ICH; March 1996.
14. International Conference on Harmonization. Guidance for industry: testing for carcinogenicity of pharmaceuticals. Geneva: ICH; July 1997; Document S1B.
15. International Conference on Harmonization. Guideline for industry: dose selection for carcinogenicity studies of pharmaceuticals. Geneva: ICH; March 1995; Document S1C.
16. International Conference on Harmonization. Guidance for industry: addendum to *Dose Selection for Carcinogenicity Studies of Pharmaceuticals:* addition of a limit dose and related notes. Geneva: ICH; March 1995; Document S1C (R).
17. The European Agency for the Evaluation of Medicinal Products Committee for Proprietary Medicinal Products. Note for guidance on carcinogenic potential. London: EMEA; July 2002.
18. International Conference on Harmonization. Guideline for industry: detection of toxicity to reproduction for medicinal products. Geneva: ICH; September 1994; Document S5A.
19. International Conference on Harmonization. Guideline for industry: detection of toxicity to reproduction for medicinal products: addendum on toxicity to male fertility. Geneva: ICH; April 1996; Document S5B.
20. International Conference on Harmonization. Guidance for industry: preclinical safety evaluation of biotechnology-derived pharmaceuticals. Geneva: ICH; July 1997; Document S6.
21. International Conference on Harmonization. Guidance for industry: safety pharmacology studies for human pharmaceuticals. Geneva: ICH; July 2001; Document S7A.
22. International Conference on Harmonization. Guideline: safety pharmacology studies for assessing the potential for delayed ventricular repolarization (QT interval prolongation) by human pharmaceuticals. Geneva: ICH; February 2002; Draft Consensus Document.
23. The European Agency for the Evaluation of Medicinal Products Committee for Proprietary Medicinal Products. Note for guidance on the pre-clinical evaluation of anticancer medicinal products. London: EMEA; July 1998.
24. 21 Code of Federal Regulations 312.21. 52 Federal Register 8831 (1981) codified at 210FR 312.21.
25. Nagamura F, Collins J, Kobayashi K, et al. Comparative review of oncology phase 1 dose escalation designs of new molecular entities. In: Proceedings of the American Society of Clinical Oncology; May 12–21, 2002; Orlando, FL. Abstract 352.
26. Simon R, Freidlin B, Rubinstein L, et al. Accelerated titration designs for phase 1 clinical trials in oncology. *J Natl Cancer Inst* 1997;89:1138–1147.
27. US Food and Drug Administration. Draft guidance for industry: IND exemptions for studies of lawfully marketed cancer drug or biological products. 67 Federal Register 17078 (2002).
28. 21 Code of Federal Regulations Part 314. 59 Federal Register 749 (1991) codified at 210FR 314.
29. 21 Code of Federal Regulations Subpart H Part 314.500: 57 Federal Register 58958 (1991) codified at 210FR 314.500.
30. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Guidance for industry: fast track drug development programs. Designation, development, and application review. September 1998.
31. Manual of Policies and Procedures 6020.3 Priority Review Policy April 22, 1996. Available at: http://www.fda.gov/cder/mapp/6020-3.
32. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research, Center for Biologics Evaluation and Research. Guidance for industry: special protocol assessment. May 2002.
33. The European Agency for the Evaluation of Medicinal Products Committee for Proprietary Medicinal Products. Note for guidance on evaluation of anticancer medicinal products in man. London: EMEA; November 2001.
34. The European Agency for the Evaluation of Medicinal Products Committee for Proprietary Medicinal Products. Accelerated evaluation of products indicated for serious disease (life-threatening or heavily disabling diseases). London: EMEA; September 2001.

General Guide for Starting Dose Selection for a Cytotoxic Agent in Cancer Patients



SPECIAL ARTICLE

Risks and Benefits of Phase 1 Oncology Trials, 1991 through 2002

Elizabeth Horstmann, B.A., Mary S. McCabe, R.N., M.A., Louise Grochow, M.D., Seiichiro Yamamoto, Ph.D., Larry Rubinstein, Ph.D., Troy Budd, Dale Shoemaker, Ph.D., Ezekiel J. Emanuel, M.D., Ph.D., and Christine Grady, R.N., Ph.D.

ABSTRACT

BACKGROUND

Previous reviews of phase 1 oncology trials reported a rate of response to treatment of 4 to 6 percent and a toxicity-related death rate of 0.5 percent. These results may not reflect the rates in current phase 1 oncology trials.

METHODS

We reviewed all nonpediatric phase 1 oncology trials sponsored by the Cancer Therapy Evaluation Program at the National Cancer Institute between 1991 and 2002. We report the rates of response to treatment, of stable disease, of grade 4 toxic events, and of treatment-related deaths.

RESULTS

We analyzed 460 trials involving 11,935 participants, all of whom were assessed for toxicity and 10,402 of whom were assessed for a response to therapy. The overall response rate (i.e., for both complete and partial responses) was 10.6 percent, with considerable variation among trials. "Classic" phase 1 trials of single investigational chemotherapeutic agents represented only 20 percent of the trials and had a response rate of 4.4 percent. Studies that included at least one anticancer agent approved by the Food and Drug Administration constituted 46.3 percent of the trials and had a response rate of 17.8. An additional 34.1 percent of participants had stable disease or a less-than-partial response. The overall rate of death due to toxic events was 0.49 percent. Of 3465 participants for whom data on patient-specific grade 4 toxic events were available, 14.3 percent had had at least one episode of grade 4 toxic events.

CONCLUSIONS

Overall response rates among phase 1 oncology trials are higher than previously reported, although they have not changed for classic phase 1 trials, and toxicity-related death rates have remained stable. Rates of response and toxicity vary, however, among the various types of phase 1 oncology trials.

From the Department of Clinical Bioethics, Warren G. Magnuson Clinical Center, National Institutes of Health, Bethesda, Md. (E.H., E.J.E., C.G.); the Memorial Sloan-Kettering Cancer Center, New York (M.S.M.); the Cancer Therapy Evaluation Program, National Cancer Institute, National Institutes of Health, Rockville, Md. (L.G., S.Y., L.R., T.B., D.S.); and the Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo (S.Y.) Address reprint requests to Dr. Grady at the Warren G. Magnuson Clinical Center, Bldg. 10, Rm. 1C118, National Institutes of Health, Bethesda, MD 20892-1156, or at cgrady@nih.gov.

N Engl J Med 2005;352:895-904.

THE ETHICAL ISSUES RAISED BY PHASE 1 oncology trials have been debated for decades.[1-6] These trials enroll patients with advanced cancer whose disease is usually refractory to available treatment in order to evaluate the safety and toxicity of new therapeutic agents, to establish the pharmacokinetic properties of those agents, and to determine a safe dose for subsequent testing.[7] Published reviews report that a tumor response occurs in 4 to 6 percent of the participants in these trials and that about 0.5 percent of participants die as the result of toxicity.[8-16] Critics of such trials cite these data when raising concerns about the poor prospect of benefit and the potential for severe harm. Some contend that the enrollment of patients with advanced disease in risky research studies with little chance of direct benefit exploits a vulnerable population.[17] The response rates of 4 to 6 percent and the toxicity-related death rate of 0.5 percent continue to be viewed as representative of phase 1 oncology trials, but these rates are based on reviews of single-agent trials. They do not take into full account the development of new types of anticancer agents, trials of combinations of agents, new trial designs, or improvements in supportive care, and they do not present a comprehensive picture of the benefits and risks associated with phase 1 trials.[18-20]

To better inform the discussion of the risks and benefits involved in phase 1 oncology trials, we reviewed studies that began between 1991 and 2002 and were sponsored by the Cancer Therapy Evaluation Program of the National Cancer Institute, the major sponsor of phase 1 oncology trials in the United States. Reflecting the full spectrum of phase 1 oncology trials, our review included trials of chemotherapeutic agents and newer, targeted agents such as antiangiogenesis factors, vaccines, and gene therapies; trials of combinations of agents, including some already approved by the Food and Drug Administration (FDA); and published and unpublished trials. To extend our understanding of the benefits and risks associated with phase 1 oncology research, data on stable disease and grade 4 toxic events are reported in addition to conventional measures of outcome.

METHODS

All nonpediatric phase 1 oncology trials sponsored by the Cancer Therapy Evaluation Program that began between 1991 and 2002 were eligible for this review, including trials that evaluated solid tumors and hematologic cancers and trials conducted at the National Institutes of Health (NIH) Clinical Center and other institutions around the United States. Excluded were phase 1–phase 2 trials, trials of radiation therapy alone, of stem-cell or bone marrow transplantation, of supportive care without anticancer agents, and of therapies for diseases other than cancer (e.g., human immunodeficiency virus disease).

The staff of the Cancer Therapy Evaluation Program plan, review, coordinate, and oversee clinical trials of investigational anticancer agents.[21] The program receives comprehensive trial data at regular intervals from investigators and actively monitors all trials through routine data submission and periodic audits. Between 1991 and 2002, data from phase 1 trials sponsored by the Cancer Therapy Evaluation Program were monitored by five different sources: the Clinical Trials Monitoring System, the Clinical Data Update System, the Annual Update System, the Quarterly Data Update, and Study Summary reports.

The Clinical Trials Monitoring System, which has been managed for the Cancer Therapy Evaluation Program by Theradex since 1979, is a database of electronically submitted case-report forms for first trials of agents in humans as well as trials of combinations of investigational new drugs and at least one FDA-approved drug that may be associated with a risk of overlapping toxic effects. Extensive data are submitted every two weeks for quality control and are maintained in a relational KnowledgeMan database (Micro Data Base Systems). Each participating institution is audited for quality assurance three times a year.

The Clinical Data Update System, managed by Capital Technology Information Systems, has received electronic data according to course of therapy and according to patient every three months since 1998. The Clinical Data Update System is generally used for late phase 1 trials of agents whose toxicity profile has been established in earlier studies. Data are maintained in a relational Oracle database. Before 1998, summary data for these trials were submitted as paper reports yearly (by the Annual Update System or by Study Summary reports), quarterly (by Quarterly Data Update), or twice a year in printed trial summaries prepared by the cooperative groups. For trials monitored by the Clinical Data Update System, the Annual Update System, Study Summary reports, and Quarterly Data Update,

each institution is audited every three years. Auditors examine the consistency of reporting, including references to source documents concerning toxic events among subjects and assessments of responses. Data reported in this article include selected variables from the database of the Cancer Therapy Evaluation Program and combine data from the program's five monitoring sources. A subgroup of 110 trials, primarily those monitored by the Annual Update System, was excluded because complete data in regard to toxicity were unavailable. None of the excluded trials were from the Clinical Trials Monitoring System's database of studies involving agents used for the first time in humans, studies involving agents filed as investigational new drugs with the FDA, or other early phase studies. The Cancer Therapy Evaluation Program provided the data on May 16, 2003.

Trials were grouped by an experienced investigator of phase 1 trials into one of six categories according to the mechanism of action of the agent or agents under investigation: cytotoxic chemotherapeutic agents, immunomodulators, receptor-transduction or signal-transduction agents (including those affecting gene reexpression), antiangiogenesis agents, gene-transfer agents, and vaccines. Each of these categories was further subdivided into four types of trials: those for single investigational agents, for multiple investigational agents, for both investigational and FDA-approved agents, and for only those agents approved by the FDA. Trials involving multiple investigational agents with different mechanisms of action were grouped according to the agent predicted to be the most toxic. Thus, any trial involving a combination of therapies that included a chemotherapeutic investigational agent was coded as a chemotherapy trial, and any trial that included an immunomodulating investigational agent but no chemotherapeutic agents was categorized as an immunomodulator trial. Trials that included both investigational and FDA-approved agents were categorized according to the mechanism of action of the investigational agent. For purposes of classification, radiation was considered an FDA-approved agent.

In cases in which the study title identified a specific disease, the study was considered disease-specific. Studies of single investigational cytotoxic chemotherapeutic agents were labeled "classic" phase 1 trials. Studies of agents being used in humans for the first time were selected from all five databases. These included the very first study of an agent conducted after the agent was filed as an investigational new drug with the FDA and trials that were initiated within seven months of the first study, before any information was available about dose-limiting toxicity from the very first trial.

Potentially beneficial effects of agents under investigation were categorized as complete response, partial response, less-than-partial response, and stable disease. Response to treatment was reported for each protocol according to guidelines of the World Health Organization (WHO),[22] the Response Evaluation Criteria in Solid Tumors,[23] or other established criteria approved by the Protocol Review Committee of the Cancer Therapy Evaluation Program. A complete response was defined as the disappearance of a tumor; a partial response as an overall 50 percent reduction in the tumor, measured as the sum of the products of the two longest diameters (according to the WHO criteria), or as an overall 30 percent reduction in tumor size, measured as the sum of the longest diameters (according to guidelines of the Response Criteria in Solid Tumors); and stable disease as neither a partial response nor progressive disease.[23] For this analysis, less-than-partial response and stable disease are combined into one category.

Toxicity was reported with the use of the Common Toxicity Criteria.[24] Protocols specified which version of these criteria were used, depending on when the protocols were initiated. All deaths reported by investigators as "possibly," "probably," or "definitely" related to treatment were considered toxicity-related deaths. Data on patient-specific grade 4 toxic events that were available from the Clinical Data Update System are reported; for the other trials, only the data on cumulative toxicity according to trial were available.

STATISTICAL ANALYSIS

Response rates, death rates, and rates of grade 4 toxic events were calculated for participants who were assessed according to trial category (i.e., therapeutic modality, single agent or combination, disease-specific or not, and first-in-human or other). Rates were calculated by dividing the total number of events (responses, deaths, or grade 4 toxic events) by the total number of patients assessed for response or toxicity. Response rates and toxicity-related death rates were also calculated for three-year intervals to evaluate trends. For the subgroup of trials monitored by the Clinical Data Update System, the percentage of patients who had grade 4

toxic events and the average number of grade 4 toxic events per affected patient were reported. Comparisons of response rates and of toxicity-related death rates — in particular, between the current sample and prior samples — were made descriptively. Calculation of statistical significance was intentionally avoided in cases where patient samples may have been divergent and hypothesis testing not prospectively defined. Statistical analyses were performed with the use of SAS software, version 8.02.

RESULTS

The sample of 460 phase 1 oncology trials sponsored by the Cancer Therapy Evaluation Program

Table 1. Rates of Response to Treatment in Phase 1 Oncology Trials.

Trial	No. of Trials	No. of Patients Assessed for Response	Rate of Response			
			Overall Response (Complete and Partial)	Complete Response	Partial Response	Stable Disease and Less-Than-Partial Response
			percent			
Total	460	10,402	10.6	3.1	7.5	34.1*
Cytotoxic chemotherapy						
One investigational agent	92	2,341	4.4	1.5	2.9	40.8
Multiple investigational agents	12	273	11.7	1.5	10.3	27.5
Combination of investigational and FDA-approved agents	88	2,251	16.4	5.6	10.8	31.3†
FDA-approved agents only	29	792	27.4	8.0	19.4	27.2†
Immunomodulator						
One investigational agent	13	203	11.3	3.0	8.4	35.5
Multiple investigational agents	28	651	6.9	2.2	4.8	22.3†
Combination of investigational and FDA-approved agents	19	392	26.0	5.6	20.4	26.7†
Receptor or signal transduction						
One investigational agent	51	1,347	3.2	0.7	2.5	39.3
Multiple investigational agents	7	81	7.4	1.2	6.2	27.2
Combination of investigational and FDA-approved agents	61	935	11.7	2.1	9.5	37.4
Antiangiogenesis						
One investigational agent	15	335	3.9	0.6	3.3	31.0
Combination of investigational and FDA-approved agents	9	135	14.8	5.2	9.6	37.0
Gene transfer						
One investigational agent	7	89	3.4	0	3.4	30.3
Combination of investigational and FDA-approved agents	1	3	0	0	0	0
Vaccine						
One investigational agent	15	265	3.4	3.0	0.4	24.9
Multiple investigational agents	7	198	1.0	1.0	0	35.4
Combination of investigational and FDA-approved agents	6	111	5.4	2.7	2.7	19.8

* For 630 of 10,402 participants, data on stable disease and less-than-partial response are not reported. The percentage was calculated with 9772 as the denominator.

† Percentages were calculated with a denominator adjusted to exclude participants for whom data on stable disease and less-than-partial response were unavailable.

that opened between 1991 and 2002 included 11,935 participants. All participants were assessed for toxicity, and 10,402 were assessed for a response (Table 1). Trials of cytotoxic chemotherapeutic agents accounted for 48.0 percent (221) of all trials and for 54.4 percent (5657) of participants assessed for response. Trials involving receptor transduction or signal transduction were the second-largest group (119 trials, or 25.9 percent), representing 22.7 percent (2363) of participants assessed for response. There were only eight trials involving gene transfer, with 92 participants (Table 1).

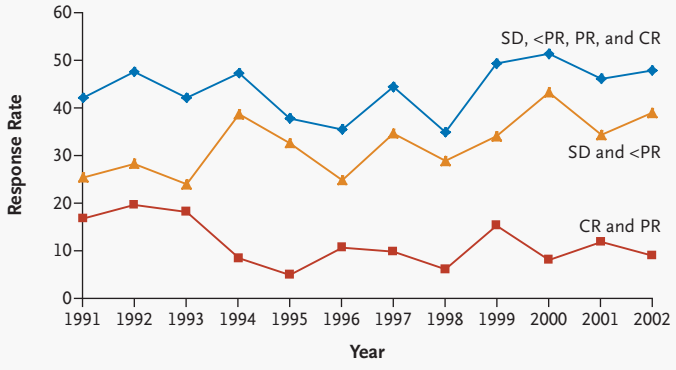


Figure 1. Response Rates According to Year.

Response to therapy was classified as complete (CR), partial (PR), less than partial (<PR), or as stable disease (SD). When the rates were grouped according to three-year periods, a downward trend was observed for complete and partial responses, but when stable disease and less-than-partial responses were taken into account, the rate remained relatively constant over time.

RESPONSE RATES

Among the trials of all types of agents, 10.6 percent of the 10,402 participants assessed for response had either a partial or a complete response to therapy. Of these, 7.5 percent had a partial response and 3.1 percent had a complete response. In addition, 34.1 percent of the participants in phase 1 trials had either stable disease or a less-than-partial response (Table 1).

Response rates varied according to the type of agent used and the characteristics of the trial (Table 1). The overall response rate was 3.0 percent among trials of vaccines and 13.6 percent among studies of immunomodulators (data not shown). Furthermore, response rates varied within categories according to the type of trial. For classic phase 1, single-agent chemotherapy studies, the overall response rate was 4.4 percent. The rate among chemotherapy studies involving more than one investigational agent was 11.7 percent; for combinations of investigational and FDA-approved agents, the rate was 16.4 percent; and for phase 1 trials including only FDA-approved chemotherapeutic agents, the rate was 27.4 percent (Table 1). A similar variation was seen in the other categories of trials (Table 1). The response rate among 3420 participants in 184 disease-specific trials was 19.3 percent; among trials that were not specific to disease, the rate was 6.3 percent.

Response rates also varied over time, with the highest rate (19.5 percent) occurring in 1992 and the lowest (5.0 percent) in 1995. When the rates were grouped according to three-year periods, a downward trend in complete and partial responses was noted (18.3 percent for 1991 to 1993 and 9.4 percent for 2000 to 2002). However, when stable disease was taken into account, the rate remained relatively constant over time (34.6 to 51.3 percent) (Fig. 1).

TOXICITY

Among the 11,935 participants in all 460 phase 1 studies, there were 58 deaths (0.49 percent) that were determined to be at least possibly related to the treatment (Table 2). Of those deaths, 18 were reported as definitely related to the treatment and 7 as probably related (for a combined toxicity-related death rate of 0.21 percent). When calculated in three-year intervals for 1991 through 2002, the toxicity-related death rate remained relatively constant (range, 0.45 to 0.61 percent). Of the 58 deaths, 43 (74.1 percent) occurred in participants in chemotherapy trials, with the highest toxicity-related death rate (0.77 percent) occurring in trials involving both investigational and FDA-approved agents (Table 2). Classic phase 1 trials of single investigational chemotherapeutic agents had a toxicity-related death rate of 0.57 percent. Thirteen deaths were reported among trials of receptor-transduction or signal-transduction agents (0.47 percent) and one death each among trials of immunomodulators (0.07 percent) and antiangiogenesis factors (0.17 percent). There were no reported deaths in phase 1 gene-transfer or vaccine studies.

In a subgroup of 168 studies that involved 3465 patients assessed for toxicity, 14.3 percent of participants had had grade 4 toxic events; an average of 1.9 grade 4 events occurred per affected patient (Table 3). On average, trials of chemotherapeutic agents were associated with the highest rate of tox-

Table 2. Deaths from Toxic Events in Phase 1 Oncology Trials.

Trial	No. of Trials	No. of Patients Assessed for Toxic Events	Deaths from Toxic Events*
			no. (%)
Total	460	11,935	58 (0.49)
Cytotoxic chemotherapy			
One investigational agent	92	2,621	15 (0.57)
Multiple investigational agents	12	305	2 (0.66)
Combination of investigational and FDA-approved agents	88	2,594	20 (0.77)
FDA-approved agents only	29	925	6 (0.65)
Immunomodulator			
One investigational agent	13	235	0
Multiple investigational agents	28	730	1 (0.14)
Combination of investigational and FDA-approved agents	19	443	0
Receptor or signal transduction			
One investigational agent	51	1,565	3 (0.19)
Multiple investigational agents	7	99	2 (2.02)
Combination of investigational and FDA-approved agents	61	1,081	8 (0.74)
Antiangiogenesis			
One investigational agent	15	402	0
Combination of investigational and FDA-approved agents	9	171	1 (0.58)
Gene transfer			
One investigational agent	7	107	0
Combination of investigational and FDA-approved agents	1	5	0
Vaccine			
One investigational agent	15	297	0
Multiple investigational agents	7	218	0
Combination of investigational and FDA-approved agents	6	137	0

* Deaths include all those reported as possibly, probably, or definitely related to the treatment.

icity, with 17.4 percent of participants experiencing at least one grade 4 toxic event; vaccine trials had the lowest rate, with no grade 4 toxic events reported (Table 3). Among all 11,935 participants assessed in the 460 studies, 5251 grade 4 toxic events were reported.

FIRST-IN-HUMAN TRIALS

Of 460 trials, 117 (25.4 percent) involving a total of 3164 participants assessed for a response to therapy were considered first-in-human trials — that is, studies designed to establish initial information on toxicity and dose for agents not previously tested in humans (Table 4). The overall response rate in these studies was 4.8 percent, as compared with 13.1 percent in the other studies. The toxicity-related death rate in first-in-human studies was 0.26 percent, as compared with 0.58 percent in studies not considered first-in-human trials. Studies of cytotoxic chemotherapeutic agents made up the largest group of first-in-human trials (36.8 percent). Of the vaccine studies sponsored by the Cancer Therapy Evaluation Program, 82.1 percent were first-in-human trials.

Table 3. Grade 4 Toxic Events in Phase 1 Oncology Trials.

Trial	No. of Trials	No. of Patients Assessed for Toxic Events	Patients with a Grade 4 Toxic Event	Average No. of Grade 4 Toxic Events per Patient
			%	
Total	168	3465	14.3	1.9
Cytotoxic chemotherapy				
One investigational agent	20	408	15.0	1.6
Multiple investigational agents	3	23	4.3	2.0
Combination of investigational and FDA-approved agents	17	475	14.5	1.8
FDA-approved agents only	3	159	34.0	2.4
Immunomodulator				
One investigational agent	2	43	2.3	1.0
Multiple investigational agents	10	207	9.7	2.2
Combination of investigational and FDA-approved agents	5	101	4.0	1.8
Receptor or signal transduction				
One investigational agent	29	839	13.0	1.7
Multiple investigational agents	6	67	19.4	2.0
Combination of investigational and FDA-approved agents	51	752	18.1	2.0
Antiangiogenesis				
One investigational agent	9	143	5.6	1.6
Combination of investigational and FDA-approved agents	6	101	17.8	1.8
Gene transfer				
One investigational agent	1	26	11.5	1.7
Combination of investigational and FDA-approved agents	1	5	0	0
Vaccine				
One investigational agent	3	20	0	0
Multiple investigational agents	2	96	0	0

TRIALS WITH FDA-APPROVED AGENTS

Overall, 213 studies (46.3 percent) included at least one FDA-approved anticancer agent. Response rates were higher in trials with FDA-approved agents than in trials without FDA-approved agents (Table 5). These studies had an overall response rate of 17.8 percent, as compared with 4.8 percent for studies not including FDA-approved anticancer agents. The toxicity-related death rate was higher (0.65 percent) than for trials that did not include FDA-approved anticancer agents (0.35 percent).

DISCUSSION

We comprehensively reviewed phase 1 oncology trials sponsored by the Cancer Therapy Evaluation Program between 1991 and 2002. The overall response rate in these trials was 10.6 percent, which is higher than previously reported, whereas the toxicity-related death rate, 0.49 percent, is similar to that of previous reports. Rates of response and of toxicity-related death among classic phase 1 trials of single chemotherapeutic agents are similar to those reported in other reviews, but classic trials account for only 22 percent of participants in this review.

Response rates in phase 1 oncology trials have been reported to be 4 to 6 percent, with toxicity-related death rates reported to be 0.5 percent or lower.[8-16] In our review, however, we found that response rates in recent phase 1 oncology trials exceeded 10 percent, with stable disease or less-than-partial re-

Table 4. Response Rates and Deaths from Toxic Events in Phase 1 Oncology Trials Involving the First Use of an Agent in Humans.

Trial	No. of Trials	No. of Patients Assessed for Response	Overall Response Rate*	No. of Patients Assessed for Toxic Events	Deaths from Toxic Events†
			%		*no. (%)*
Total					
First use of an agent in humans	117	3164	4.8	3498	9 (0.26)
All other trials	343	7238	13.1	8437	49 (0.58)
Cytotoxic chemotherapy					
First use of an agent in humans	43	1298	5.0	1422	7 (0.49)
All other trials	178	4359	15.0	5023	36 (0.72)
Immunomodulator					
First use of an agent in humans	16	404	7.4	431	1 (0.23)
All other trials	44	842	16.6	977	0
Receptor or signal transduction					
First use of an agent in humans	27	742	3.8	853	1 (0.12)
All other trials	92	1621	8.0	1892	12 (0.63)
Antiangiogenesis					
First use of an agent in humans	8	200	7.0	228	0
All other trials	16	270	7.0	345	1 (0.29)
Gene transfer					
First use of an agent in humans	0	0	0	0	0
All other trials	8	92	3.3	112	0
Vaccine					
First use of an agent in humans	23	520	3.1	564	0
All other trials	5	54	1.9	88	0

* The overall response rate includes both complete and partial responses.
† Deaths include all those reported as possibly, probably, or definitely related to the treatment.

sponse having been achieved in an additional 34.1 percent of participants. Rates of toxicity-related death have not increased over time, and more than 85 percent of participants had no grade 4 toxic events. As compared with other reviews, these data suggest that participants may benefit more from current phase 1 oncology trials than previously believed.

A recent review of single-agent trials showed that there was a decrease in tumor-response rates over time,[13] which was attributed to the use of newer, more specific agents and changes in trial design. In our review, response rates per year varied without a clear pattern. When these rates were grouped in three-year intervals, there was a decrease in complete or partial responses from 1991 to 2002 but an increase in rates of stable disease. Little change in the benefit to participants over time was seen when response rates were grouped with stable disease.

In our view, it is inaccurate to refer to phase 1 oncology studies as if they are all similar to one another. Nearly half of the trials we studied included at least one FDA-approved agent, and less than half included chemotherapeutic agents. Different types of phase 1 oncology studies are associated with very different response rates. For instance, the response rate among patients who were treated with immunomodulators was 13.6 percent, yet the rate was just 3.0 percent for patients treated with vaccines. Trials that included one or more FDA-approved anticancer agents showed higher response rates than did those involving only investigational agents. For these reasons, it may be misleading to summarize phase 1 oncology trials with the use of a single response rate.

Risk, as measured by toxicity-related death rates and grade 4 toxic events, also varies according to the type of trial. The average toxicity-related death rate for trials of cytotoxic chemotherapeutic agents was 0.67 percent but just 0.07 percent for those in-

Table 5. Response Rates and Deaths from Toxic Events in Phase 1 Oncology Trials, According to Whether FDA-Approved Agents Were Used.

Trial	No. of Trials	No. of Patients Assessed for Response	Overall Response Rate*	No. of Patients Assessed for Toxic Events	Deaths from Toxic Events†
			%		*no. (%)*
Single investigational agent	193	4580	4.2	5227	18 (0.34)
Multiple investigational agents	54	1203	7.1	1352	5 (0.37)
Combination of investigational and FDA-approved agents	184	3827	15.8	4431	29 (0.65)
FDA-approved agents only	29	792	27.4	925	6 (0.65)

* The overall response rate includes both complete and partial responses.
† Deaths include all those reported as possibly, probably, or definitely related to the treatment.

volving immunomodulators, and no toxicity-related deaths were reported in gene-transfer or vaccine trials. Grade 4 toxic events were more common in chemotherapy trials, especially those involving multiple agents, than in all other trials. Trials of FDA-approved drugs, which evaluated the safety of higher doses or combinations of drugs, appeared to be associated with the highest rates of toxicity (a death rate from toxic events of 0.65 percent, vs. 0.35 percent for other trials) but also had the highest overall response rate (17.8 percent, vs. 4.8 percent for other trials). Overall, newer, nonchemotherapeutic agents are associated with lower rates of toxic events.

Classic phase 1 studies of single investigational chemotherapeutic agents, which were the only trials included in previous reviews, showed an overall response rate of 4.4 percent and a toxicity-related death rate of 0.57 percent. These rates are almost identical to those previously reported.[8-16] In this study of trials sponsored by the Cancer Therapy Evaluation Program and initiated between 1991 and 2002, classic phase 1 trials accounted for only 22 percent of all participants. Similarly, the testing of investigational agents never before studied in humans is commonly thought of as a defining characteristic of phase 1 oncology trials. In our review, these first-in-human studies represented less than a quarter of phase 1 studies and enrolled less than a third of participants. Response rates, but also toxicity-related death rates, are lower in studies that test agents for the first time in humans than in those that do not test agents for the first time.

When the risks and benefits associated with phase 1 oncology trials are weighed, factors other than response rates and toxicity should be taken into account. Investigational treatments may have clinically meaningful benefits — reduced pain, increased appetite, energy, and activity, weight gain, reduced fatigue, or increased ability to perform daily activities.[20,25,26] Some of these benefits might accrue from research participation itself; for some persons, contributing to research and potentially helping future cancer patients may also be an important benefit.[27] At the same time, participation in research may involve additional burdens: multiple visits or long hours at the clinic, unpleasant procedures, and the possible financial costs associated with participation in research studies.[28]

This study has several limitations. First, our data are derived only from trials sponsored by the Cancer Therapy Evaluation Program. Although the program is a major sponsor of phase 1 oncology trials in the United States[29] and the use of data from the program avoids publication bias, any differences that might be found in the phase 1 trials with other sponsors have not been captured. It is possible that the response rates associated with trials of promising agents sponsored by pharmaceutical companies could be higher than those reported here. Second, for trials involving gene transfer, the findings should be interpreted with caution because of the small number of trials and the possibility that outliers influenced the data. Finally, our reporting of grade 4 toxic events is limited. Patient-specific data on grade 4 toxic events came from one monitoring source, which, although it includes some first-in-human trials, is generally used to monitor later phase 1 studies and may not be entirely representative of phase 1 oncology studies. Moreover, the data on grade 4 toxic events are reported without distinguishing among the types of toxic events.

Since not all toxic events have similar medical consequences, evaluation of the risks in phase 1 trials should include both the types and the frequency of events experienced by participants.

In conclusion, reliance on a single estimate of the response rate or the toxicity-related death rate for phase 1 oncology trials is misleading, since rates of response and toxicity vary according to the type of trial. Potential participants and their families, oncologists, investigators, members of institutional review boards, ethicists, and others interested in weighing the risks and benefits of phase 1 studies and making decisions about their acceptability should be aware of the complexity and variety of such trials, know the details about the trial they are considering, and carefully evaluate all relevant risks and benefits.

Supported in part by the Department of Clinical Bioethics, National Institutes of Health, and by a grant (to Dr. Yamamoto) from the Ministry of Health, Labor, and Welfare of Japan.

Ms. McCabe reports having received consulting fees from the Lance Armstrong Foundation and the American Society of Clinical Oncology. Dr. Emanuel reports having received lecture fees from Merck and owning stock or equity in Biogen, Neurobiological Technologies, and Titan Pharmaceuticals.

The views expressed in this article are those of the authors and do not represent the views or policies of the Department of Health and Human Services or the National Institutes of Health.

We are indebted to the countless investigators and the 11,935 patients with cancer who made this review possible by participating in the oncology trials of the Cancer Therapy Evaluation Program; and to Manish Agrawal and Michaele Christian for their contributions to this article.

REFERENCES

1. Lipsett MB. On the nature and ethics of phase I clinical trials of cancer chemotherapies. JAMA 1982;248:941-2.
2. Kodish E, Stocking C, Ratain MJ, Kohrman A, Siegler M. Ethical issues in phase I oncology research: a comparison of investigators and institutional review board chairpersons. J Clin Oncol 1992;10:1810-6.
3. Agrawal M, Emanuel EJ. Ethics of phase 1 oncology studies: reexamining the arguments and data. JAMA 2003;290:1075-82.
4. Miller M. Phase I cancer trials: a collusion of misunderstanding. Hastings Cent Rep 2000;30:34-43.
5. *Idem.* Phase I cancer trials: a crucible of competing priorities. Int Anesthesiol Clin 2001;39:13-33.
6. Emanuel EJ. A phase I trial on the ethics of phase I trials. J Clin Oncol 1995;13:1049-51.
7. Markman M. The needs of science vs the needs of patients: ethical concerns in cancer clinical trials. Cleve Clin J Med 2003;70:1008-9, 1013-4, 1016.
8. Itoh K, Sasaki Y, Miyata Y, et al. Therapeutic response and potential pitfalls in phase I clinical trials of anticancer agents conducted in Japan. Cancer Chemother Pharmacol 1994;34:451-4.
9. Estey E, Hoth D, Simon R, Marsoni S, Leyland-Jones B, Wittes R. Therapeutic response in phase I trials of antineoplastic agents. Cancer Treat Rep 1986;70:1105-15.
10. Decoster G, Stein G, Holdener EE. Responses and toxic deaths in phase I clinical trials. Ann Oncol 1990;1:175-81.
11. Bachelot T, Ray-Coquard I, Catimel G, et al. Multivariable analysis of prognostic factors for toxicity and survival for patients enrolled in phase I clinical trials. Ann Oncol 2000;11:151-6.
12. Han C, Braybrooke JP, Deplanque G, et al. Comparison of prognostic factors in patients in phase I trials of cytotoxic drugs vs new noncytotoxic agents. Br J Cancer 2003;89:1166-71.
13. Roberts TG Jr, Goulart BH, Squitieri L, et al. Trends in the risks and benefits to patients with cancer participating in phase 1 clinical trials. JAMA 2004;292:2130-40.
14. Sekine I, Yamamoto N, Kunitoh H, et al. Relationship between objective responses in phase I trials and potential efficacy of nonspecific cytotoxic investigational new drugs. Ann Oncol 2002;13:1300-6.
15. Smith TL, Lee JJ, Kantarjian HM, Legha SS, Raber MN. Design and results of phase I cancer clinical trials: three-year experience at M.D. Anderson Cancer Center. J Clin Oncol 1996;14:287-95.
16. Von Hoff DD, Turner J. Response rates, duration of response, and dose response effects in phase I studies of antineoplastics. Invest New Drugs 1991;9:115-22.
17. Annas GJ. The changing landscape of human experimentation: Nuremberg, Helsinki, and beyond. Health Matrix Clevel 1992;2:119-40.
18. Dent SF, Eisenhauer EA. Phase I trial design: are new methodologies being put into practice? Ann Oncol 1996;7:561-6.
19. Eisenhauer EA, O'Dwyer PJ, Christian M, Humphrey JS. Phase I clinical trial design in cancer drug development. J Clin Oncol 2000;18:684-92.
20. Chen EX, Tannock IF. Risks and benefits of phase 1 clinical trials evaluating new anticancer agents: a case for more innovation. JAMA 2004;292:2150-1.
21. Ansher SS, Scharf R. The Cancer Therapy Evaluation Program (CTEP) at the National Cancer Institute: industry collaborations in new agent development. Ann N Y Acad Sci 2001;949:333-40.
22. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207-14.
23. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors: European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 2000;92:205-16.
24. Cancer Therapy Evaluation Program. Common Toxicity Criteria (CTC), version 2.0. Bethesda, Md.: National Cancer Institute, 1999. (Accessed February 14, 2005, at http://ctep.cancer.gov/reporting/ctc.html.)
25. Markman M. Partial responses to antineoplastic drug therapy: what do they mean? J Cancer Res Clin Oncol 1992;119:5-6.
26. Shanafelt TD, Loprinzi C, Marks R, Novotny P, Sloan J. Are chemotherapy response rates related to treatment-induced survival prolongations in patients with advanced cancer? J Clin Oncol 2004;22:1966-74.
27. Daugherty CK, Siegler M, Ratain MJ, Zimmer G. Learning from our patients: one participant's impact on clinical trial research and informed consent. Ann Intern Med 1997;126:892-7.
28. Goldman DP, Berry SH, McCabe MS, et al. Incremental treatment costs in National Cancer Institute-sponsored clinical trials. JAMA 2003;289:2970-7.
29. Takimoto CH. Anticancer drug development at the US National Cancer Institute. Cancer Chemother Pharmacol 2003;52:Suppl 1:S29-S33.

SPECIAL ARTICLE

New Guidelines to Evaluate the Response to Treatment in Solid Tumors

Patrick Therasse, Susan G. Arbuck, Elizabeth A. Eisenhauer, Jantien Wanders, Richard S. Kaplan, Larry Rubinstein, Jaap Verweij, Martine Van Glabbeke, Allan T. van Oosterom, Michaele C. Christian, Steve G. Gwyther

Anticancer cytotoxic agents go through a process by which their antitumor activity—on the basis of the amount of tumor shrinkage they could generate—has been investigated. In the late 1970s, the International Union Against Cancer and the World Health Organization introduced specific criteria for the codification of tumor response evaluation. In 1994, several organizations involved in clinical research combined forces to tackle the review of these criteria on the basis of the experience and knowledge acquired since then. After several years of intensive discussions, a new set of guidelines is ready that will supersede the former criteria. In parallel to this initiative, one of the participating groups developed a model by which response rates could be derived from unidimensional measurement of tumor lesions instead of the usual bidimensional approach. This new concept has been largely validated by the Response Evaluation Criteria in Solid Tumors Group and integrated into the present guidelines. This special article also provides some philosophic background to clarify the various purposes of response evaluation. It proposes a model by which a combined assessment of all existing lesions, characterized by target lesions (to be measured) and nontarget lesions, is used to extrapolate an overall response to treatment. Methods of assessing tumor lesions are better codified, briefly within the guidelines and in more detail in Appendix I. All other aspects of response evaluation have been discussed, reviewed, and amended whenever appropriate. [J Natl Cancer Inst 2000; 92:205–16]

A. Preamble

Early attempts to define the objective response of a tumor to an anticancer agent were made in the early 1960s *(1,2).* In the mid- to late 1970s, the definitions of objective tumor response were widely disseminated and adopted when it became apparent that a common language would be necessary to report the results of cancer treatment in a consistent manner.

The World Health Organization (WHO) definitions published in the 1979 *WHO Handbook (3)* and by Miller et al. *(4)* in 1981 have been the criteria most commonly used by investigators around the globe. However, some problems have developed with the use of WHO criteria: 1) The methods for integrating into response assessments the change in size of measurable and "evaluable" lesions as defined by WHO vary among research groups, 2) the minimum lesion size and number of lesions to be recorded also vary, 3) the definitions of progressive disease are related to change in a single lesion by some and to a change in the overall tumor load (sum of the measurements of all lesions) by others, and 4) the arrival of new technologies (computed tomography [CT] and magnetic resonance imaging [MRI]) has led to some confusion about how to integrate three-dimensional measures into response assessment.

These issues and others have led to a number of different modifications or clarifications to the WHO criteria, resulting in a situation where response criteria are no longer comparable among research organizations—the very circumstance that the WHO publication had set out to avoid. This situation led to an initiative undertaken by representatives of several research groups to review the response definitions in use and to create a revision of the WHO criteria that, as far as possible, addressed areas of conflict and inconsistency.

In so doing, a number of principles were identified:

1) Despite the fact that "novel" therapies are being developed that may work by mechanisms unlikely to cause tumor regression, there remains an important need to continue to describe objective change in tumor size in solid tumors for the foreseeable future. Thus, the four categories of complete response, partial response, stable disease, and progressive disease, as originally categorized in the *WHO Handbook (3),* should be retained in any new revision.
2) Because of the need to retain some ability to compare favorable results of future therapies with those currently available, it was agreed that no major discrepancy in the meaning and the concept of partial response should exist between the old and the new guidelines, although measurement criteria would be different.
3) In some institutions, the technology now exists to determine

Affiliations of authors: P. Therasse, J. Verweij, M. Van Glabbeke, A. T. van Oosterom, European Organization for Research and Treatment of Cancer, Brussels, Belgium; S. G. Arbuck, R. S. Kaplan, L. Rubinstein, M. C. Christian, National Cancer Institute, Bethesda, MD; E. A. Eisenhauer, National Cancer Institute of Canada Clinical Trials Group, Kingston, ON, Canada; J. Wanders, New Drug Development Office Oncology, Amsterdam, The Netherlands; S. G. Gwyther, East Surrey Healthcare National Health Service Trust, Redhill, U.K.

Correspondence to: Patrick Therasse, M.D., European Organization for Research and Treatment of Cancer Data Center, Avenue Mounier 83/11, 1200 Brussels, Belgium (e-mail: pth@eortc.be).

See "Note" following "References."

changes in tumor volume or changes in tumor metabolism that may herald shrinkage. However, these techniques are not yet widely available, and many have not been validated. Furthermore, it was recognized that the utility of response criteria to date had not been related to precision of measurement. The definition of a partial response, in particular, is an arbitrary convention—there is no inherent meaning for an individual patient of a 50% decrease in overall tumor load. It was not thought that increased precision of measurement of tumor volume was an important goal for its own sake. Rather, standardization and simplification of methodology were desirable. Nevertheless, the guidelines proposed in this document are not meant to discourage the development of new tools that may provide more reliable surrogate end points than objective tumor response for predicting a potential therapeutic benefit for cancer patients.

4) Concerns regarding the ease with which a patient may be considered mistakenly to have disease progression by the current WHO criteria (primarily because of measurement error) have already led some groups such as the Southwest Oncology Group to adopt criteria that require a greater increase in size of the tumor to consider a patient to have progressive disease *(5)*. These concerns have led to a similar change within these revised WHO criteria (*see* Appendix II).
5) These criteria have not addressed several other areas of recent concern, but it is anticipated that this process will continue and the following will be considered in the future:
   - Measures of antitumor activity, other than tumor shrinkage, that may appropriately allow investigation of cytostatic agents in phase II trials;
   - Definitions of serum marker response and recommended methodology for their validation; and
   - Specific tumors or anatomic sites presenting unique complexities.

B. Background

These guidelines are the result of a large, international collaboration. In 1994, the European Organization for Research and Treatment of Cancer (EORTC), the National Cancer Institute (NCI) of the United States, and the National Cancer Institute of Canada Clinical Trials Group set up a task force (*see* Appendix III) with the main objective of reviewing the existing sets of criteria used to evaluate response to treatment in solid tumors. After 3 years of regular meetings and exchange of ideas within the task force, a draft revised version of the WHO criteria was produced and widely circulated (*see* Appendix IV). Comments received (response rate, 95%) were compiled and discussed within the task force before a second version of the document integrating relevant comments was issued. This second version of the document was again circulated to external reviewers who were also invited to participate in a consensus meeting (on behalf of the organization that they represented) to discuss and finalize unresolved problems (October 1998). The list of participants to this consensus meeting is shown in Appendix IV and included representatives from academia, industry, and regulatory authorities. Following the recommendations discussed during the consensus meeting, a third version of the document was produced, presented publicly to the scientific community (American Society for Clinical Oncology, 1999), and submitted to the *Journal of the National Cancer Institute* in June 1999 for official publication.

Data from collaborative studies, including more than 4000 patients assessed for tumor response, support the simplification of response evaluation through the use of unidimensional measurements and the sum of the longest diameters instead of the conventional method using two measurements and the sum of the products. The results of the different retrospective analyses (comparing both approaches) performed by use of these different databases are described in Appendix V. This new approach, which has been implemented in the following guidelines, is based on the model proposed by James et al. *(6)*.

C. Response Evaluation Criteria in Solid Tumors (RECIST) Guidelines

1. Introduction

The introduction explores the definitions, assumptions, and purposes of tumor response criteria. Below, guidelines that are offered may lead to more uniform reporting of outcomes of clinical trials. Note that, although single investigational agents are discussed, the principles are the same for drug combinations, noninvestigational agents, or approaches that do not involve drugs.

Tumor response associated with the administration of anticancer agents can be evaluated for at least three important purposes that are conceptually distinct:

- Tumor response as a prospective end point in early clinical trials. In this situation, objective tumor response is employed to determine whether the agent/regimen demonstrates sufficiently encouraging results to warrant further testing. These trials are typically phase II trials of investigational agents/regimens (*see* section 1.2), and it is for use in this precise context that these guidelines have been developed.
- Tumor response as a prospective end point in more definitive clinical trials designed to provide an estimate of benefit for a specific cohort of patients. These trials are often randomized comparative trials or single-arm comparisons of combinations of agents with historical control subjects. In this setting, objective tumor response is used as a surrogate end point for other measures of clinical benefit, including time to event (death or disease progression) and symptom control (*see* section 1.3).
- Tumor response as a guide for the clinician and patient or study subject in decisions about continuation of current therapy. This purpose is applicable both to clinical trials and to routine practice (*see* section 1.1), but use in the context of decisions regarding continuation of therapy is not the primary focus of this document.

However, in day-to-day usage, the distinction among these uses of the term "tumor response" can easily be missed, unless an effort is made to be explicit. When these differences are ignored, inappropriate methodology may be used and incorrect conclusions may result.

1.1. Response Outcomes in Daily Clinical Practice of Oncology

The evaluation of tumor response in the daily clinical practice of oncology may not be performed according to predefined criteria. It may, rather, be based on a subjective medical judgment that results from clinical and laboratory data that are used to assess the treatment benefit for the patient. The defined criteria

developed further in this document are not necessarily applicable or complete in such a context. It might be appropriate to make a distinction between "clinical improvement" and "objective tumor response" in routine patient management outside the context of a clinical trial.

1.2. Response Outcomes in Uncontrolled Trials as a Guide to Further Testing of a New Therapy

"Observed response rate" is often employed in single-arm studies as a "screen" for new anticancer agents that warrant further testing. Related outcomes, such as response duration or proportion of patients with complete responses, are sometimes employed in a similar fashion. The utilization of a response rate in this way is not encumbered by an implied assumption about the therapeutic benefit of such responses but rather implies some degree of biologic antitumor activity of the investigated agent.

For certain types of agents (i.e., cytotoxic drugs and hormones), experience has demonstrated that objective antitumor responses observed at a rate higher than would have been expected to occur spontaneously can be useful in selecting anticancer agents for further study. Some agents selected in this way have eventually proven to be clinically useful. Furthermore, criteria for "screening" new agents in this way can be modified by accumulated experience and eventually validated in terms of the efficiency by which agents so screened are shown to be of clinical value by later, more definitive, trials.

In most circumstances, however, a new agent achieving a response rate determined *a priori* to be sufficiently interesting to warrant further testing may not prove to be an effective treatment for the studied disease in subsequent randomized phase III trials. Random variables and selection biases, both known and unknown, can have an overwhelming effect in small, uncontrolled trials. These trials are an efficient and economic step for initial evaluation of the activity of a new agent or combination in a given disease setting. However, many such trials are performed, and the proportion that will provide false-positive results is necessarily substantial. In many circumstances, it would be appropriate to perform a second small confirmatory trial before initiating large resource-intensive phase III trials.

Sometimes, several new therapeutic approaches are studied in a randomized phase II trial. The purpose of randomization in this setting, as in phase III studies, is to minimize the impact of random imbalances in prognostic variables. However, randomized phase II studies are, by definition, not intended to provide an adequately powered comparison between arms (regimens). Rather, the goal is simply to identify one or more arms for further testing, and the sample size is chosen so to provide reasonable confidence that a truly inferior arm is not likely to be selected. Therefore, reporting the results of such randomized phase II trials should not imply statistical comparisons between treatment arms.

1.3. Response Outcomes in Clinical Trials as a Surrogate for Palliative Effect

1.3.1. Use in nonrandomized clinical trials. The only circumstance in which objective responses in a nonrandomized trial can permit a tentative assumption of a palliative effect (i.e., beyond a purely clinical measure of benefit) is when there is an actual or implied comparison with historical series of similar patients. This assumption is strongest when the prospectively determined statistical analysis plan provides for matching of relevant prognostic variables between case subjects and a defined series of control subjects. Otherwise, there must be, at the very least, prospectively determined statistical criteria that provide a very strong justification for assumptions about the response rate that would have been expected in the appropriate "control" population (untreated or treated with conventional therapy, as fits the clinical setting). However, even under these circumstances, a high rate of observed objective response does not constitute proof or confirmation of clinical therapeutic benefit. Because of unavoidable and nonquantifiable biases inherent in nonrandomized trials, proof of benefit still requires eventual confirmation in a prospectively randomized, controlled trial of adequate size. The appropriate end points of therapeutic benefit for such a trial are survival, progression-free survival, or symptom control (including quality of life).

1.3.2. Use in randomized trials. Even in the context of prospectively randomized phase III comparative trials, "observed response rate" should not be the sole, or major, end point. The trial should be large enough that differences in response rate can be validated by association with more definitive end points reflecting therapeutic benefit, such as survival, progression-free survival, reduction in symptoms, or improvement (or maintenance) of quality of life.

2. Measurability of Tumor Lesions at Baseline

2.1. Definitions

At baseline, tumor lesions will be categorized as follows: measurable (lesions that can be accurately measured in at least one dimension [longest diameter to be recorded] as $\geqslant$20 mm with conventional techniques or as $\geqslant$10 mm with spiral CT scan [*see* section 2.2]) or nonmeasurable (all other lesions, including small lesions [longest diameter <20 mm with conventional techniques or <10 mm with spiral CT scan] and truly nonmeasurable lesions).

The term "evaluable" in reference to measurability is not recommended and will not be used because it does not provide additional meaning or accuracy.

All measurements should be recorded in metric notation by use of a ruler or calipers. All baseline evaluations should be performed as closely as possible to the beginning of treatment and never more than 4 weeks before the beginning of treatment.

Lesions considered to be truly nonmeasurable include the following: bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusion, inflammatory breast disease, lymphangitis cutis/pulmonis, abdominal masses that are not confirmed and followed by imaging techniques, and cystic lesions.

(*Note:* Tumor lesions that are situated in a previously irradiated area might or might not be considered measurable, and the conditions under which such lesions should be considered must be defined in the protocol when appropriate.)

2.2. Specifications by Methods of Measurements

The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up. Imaging-based evaluation is preferred to evaluation by clinical examination when both methods have been used to assess the antitumor effect of a treatment.

2.2.1. Clinical examination. Clinically detected lesions will only be considered measurable when they are superficial (e.g., skin nodules and palpable lymph nodes). For the case of skin lesions, documentation by color photography—including a ruler to estimate the size of the lesion—is recommended.

2.2.2. Chest x-ray. Lesions on chest x-ray are acceptable as measurable lesions when they are clearly defined and surrounded by aerated lung. However, CT is preferable. More details concerning the use of this method of assessment for objective tumor response evaluation are provided in Appendix I.

2.2.3. CT and MRI. CT and MRI are the best currently available and most reproducible methods for measuring target lesions selected for response assessment. Conventional CT and MRI should be performed with contiguous cuts of 10 mm or less in slice thickness. Spiral CT should be performed by use of a 5-mm contiguous reconstruction algorithm; this specification applies to the tumors of the chest, abdomen, and pelvis, while head and neck tumors and those of the extremities usually require specific protocols. More details concerning the use of these methods of assessment for objective tumor response evaluation are provided in Appendix I.

2.2.4. Ultrasound. When the primary end point of the study is objective response evaluation, ultrasound should not be used to measure tumor lesions that are clinically not easily accessible. It may be used as a possible alternative to clinical measurements for superficial palpable lymph nodes, subcutaneous lesions, and thyroid nodules. Ultrasound might also be useful to confirm the complete disappearance of superficial lesions usually assessed by clinical examination. Justifications for not using ultrasound to measure tumor lesions for objective response evaluation are provided in Appendix I.

2.2.5. Endoscopy and laparoscopy. The utilization of these techniques for objective tumor evaluation has not yet been fully or widely validated. Their uses in this specific context require sophisticated equipment and a high level of expertise that may be available only in some centers. Therefore, utilization of such techniques for objective tumor response should be restricted to validation purposes in specialized centers. However, such techniques can be useful in confirming complete histopathologic response when biopsy specimens are obtained.

2.2.6. Tumor markers. Tumor markers alone cannot be used to assess response. However, if markers are initially above the upper normal limit, they must return to normal levels for a patient to be considered in complete clinical response when all tumor lesions have disappeared. Specific additional criteria for standardized usage of prostate-specific antigen and CA (cancer antigen) 125 response in support of clinical trials are being validated.

2.2.7. Cytology and histology. Cytologic and histologic techniques can be used to differentiate between partial response and complete response in rare cases (e.g., after treatment to differentiate between residual benign lesions and residual malignant lesions in tumor types such as germ cell tumors). Cytologic confirmation of the neoplastic nature of any effusion that appears or worsens during treatment is required when the measurable tumor has met criteria for response or stable disease. Under such circumstances, the cytologic examination of the fluid collected will permit differentiation between response or stable disease (an effusion may be a side effect of the treatment) and progressive disease (if the neoplastic origin of the fluid is confirmed). New techniques to better establish objective tumor response will be integrated into these criteria when they are fully validated to be used in the context of tumor response evaluation.

3. Tumor Response Evaluation

3.1. Baseline Evaluation

3.1.1. Assessment of overall tumor burden and measurable disease. To assess objective response, it is necessary to estimate the overall tumor burden at baseline to which subsequent measurements will be compared. Only patients with measurable disease at baseline should be included in protocols where objective tumor response is the primary end point. Measurable disease is defined by the presence of at least one measurable lesion (as defined in section 2.1). If the measurable disease is restricted to a solitary lesion, its neoplastic nature should be confirmed by cytology/histology.

3.1.2. Baseline documentation of "target" and "nontarget" lesions. All measurable lesions up to a maximum of five lesions per organ and 10 lesions in total, representative of all involved organs, should be identified as target lesions and recorded and measured at baseline. Target lesions should be selected on the basis of their size (those with the longest diameter) and their suitability for accurate repeated measurements (either by imaging techniques or clinically). A sum of the longest diameter for all target lesions will be calculated and reported as the baseline sum longest diameter. The baseline sum longest diameter will be used as the reference by which to characterize the objective tumor response.

All other lesions (or sites of disease) should be identified as nontarget lesions and should also be recorded at baseline. Measurements of these lesions are not required, but the presence or absence of each should be noted throughout follow-up.

3.2. Response Criteria

3.2.1. Evaluation of target lesions. This section provides the definitions of the criteria used to determine objective tumor response for target lesions. The criteria have been adapted from the original *WHO Handbook (3),* taking into account the measurement of the longest diameter only for all target lesions: complete response—the disappearance of all target lesions; partial response—at least a 30% decrease in the sum of the longest diameter of target lesions, taking as reference the baseline sum longest diameter; progressive disease—at least a 20% increase in the sum of the longest diameter of target lesions, taking as reference the smallest sum longest diameter recorded since the treatment started or the appearance of one or more new lesions; stable disease—neither sufficient shrinkage to qualify for partial response nor sufficient increase to qualify for progressive disease, taking as reference the smallest sum longest diameter since the treatment started.

3.2.2. Evaluation of nontarget lesions. This section provides the definitions of the criteria used to determine the objective tumor response for nontarget lesions: complete response—the disappearance of all nontarget lesions and normalization of tumor marker level; incomplete response/stable disease—the persistence of one or more nontarget lesion(s) and/or the maintenance of tumor marker level above the normal limits; and progressive disease—the appearance of one or more new lesions and/or unequivocal progression of existing nontarget lesions *(1).*

(*Note:* Although a clear progression of "nontarget" lesions only is exceptional, in such circumstances, the opinion of the

treating physician should prevail and the progression status should be confirmed later by the review panel [or study chair]).

3.2.3. Evaluation of best overall response. The best overall response is the best response recorded from the start of treatment until disease progression/recurrence (taking as reference for progressive disease the smallest measurements recorded since the treatment started). In general, the patient's best response assignment will depend on the achievement of both measurement and confirmation criteria (*see* section 3.3.1). Table 1 provides overall responses for all possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new lesions.

(*Notes:*

- Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be classified as having "symptomatic deterioration." Every effort should be made to document the objective disease progression, even after discontinuation of treatment.
- Conditions that may define early progression, early death, and inevaluability are study specific and should be clearly defined in each protocol (depending on treatment duration and treatment periodicity).
- In some circumstances, it may be difficult to distinguish residual disease from normal tissue. When the evaluation of complete response depends on this determination, it is recommended that the residual lesion be investigated (fine-needle aspiration/biopsy) before confirming the complete response status.)

3.2.4. Frequency of tumor re-evaluation. Frequency of tumor re-evaluation while on treatment should be protocol specific and adapted to the type and schedule of treatment. However, in the context of phase II studies where the beneficial effect of therapy is not known, follow-up of every other cycle (i.e., 6–8 weeks) seems a reasonable norm. Smaller or greater time intervals than these could be justified in specific regimens or circumstances.

After the end of the treatment, the need for repetitive tumor evaluations depends on whether the phase II trial has, as a goal, the response rate or the time to an event (disease progression/death). If time to an event is the main end point of the study, then routine re-evaluation is warranted of those patients who went off the study for reasons other than the expected event at frequencies to be determined by the protocol. Intervals between evaluations twice as long as on study are often used, but no strict rule can be made.

Table 1. Overall responses for all possible combinations of tumor responses in target and nontarget lesions with or without the appearance of new lesions*

Target lesions	Nontarget lesions	New lesions	Overall response
CR	CR	No	CR
CR	Incomplete response/SD	No	PR
PR	Non-PD	No	PR
SD	Non-PD	No	SD
PD	Any	Yes or no	PD
Any	PD	Yes or no	PD
Any	Any	Yes	PD

*CR = complete response; PR = partial response; SD = stable disease; and PD = progressive disease. *See* text for more details.

3.3. Confirmatory Measurement/Duration of Response

3.3.1. Confirmation. The main goal of confirmation of objective response in clinical trials is to avoid overestimating the response rate observed. This aspect of response evaluation is particularly important in nonrandomized trials where response is the primary end point. In this setting, to be assigned a status of partial response or complete response, changes in tumor measurements must be confirmed by repeat assessments that should be performed no less than 4 weeks after the criteria for response are first met. Longer intervals as determined by the study protocol may also be appropriate.

In the case of stable disease, measurements must have met the stable disease criteria at least once after study entry at a minimum interval (in general, not less than 6–8 weeks) that is defined in the study protocol (*see* section 3.3.3).

(*Note:* Repeat studies to confirm changes in tumor size may not always be feasible or may not be part of the standard practice in protocols where progression-free survival and overall survival are the key end points. In such cases, patients will not have "confirmed response." This distinction should be made clear when reporting the outcome of such studies.)

3.3.2. Duration of overall response. The duration of overall response is measured from the time that measurement criteria are met for complete response or partial response (whichever status is recorded first) until the first date that recurrent or progressive disease is objectively documented (taking as reference for progressive disease the smallest measurements recorded since the treatment started). The duration of overall complete response is measured from the time measurement criteria are first met for complete response until the first date that recurrent disease is objectively documented.

3.3.3. Duration of stable disease. Stable disease is measured from the start of the treatment until the criteria for disease progression is met (taking as reference the smallest measurements recorded since the treatment started). The clinical relevance of the duration of stable disease varies for different tumor types and grades. Therefore, it is highly recommended that the protocol specify the minimal time interval required between two measurements for determination of stable disease. This time interval should take into account the expected clinical benefit that such a status may bring to the population under study.

(*Note:* The duration of response or stable disease as well as the progression-free survival are influenced by the frequency of follow-up after baseline evaluation. It is not in the scope of this guideline to define a standard follow-up frequency that should take into account many parameters, including disease types and stages, treatment periodicity, and standard practice. However, these limitations to the precision of the measured end point should be taken into account if comparisons among trials are to be made.)

3.4. Progression-Free Survival/Time to Progression

This document focuses primarily on the use of objective response end points. In some circumstances (e.g., brain tumors or investigation of noncytoreductive anticancer agents), response evaluation may not be the optimal method to assess the potential anticancer activity of new agents/regimens. In such cases, progression-free survival/time to progression can be considered valuable alternatives to provide an initial estimate of biologic effect of new agents that may work by a noncytotoxic mecha-

nism. It is clear though that, in an uncontrolled trial proposing to utilize progession-free survival/time to progression, it will be necessary to document with care the basis for estimating what magnitude of progression-free survival/time to progression would be expected in the absence of a treatment effect. It is also recommended that the analysis be quite conservative in recognition of the likelihood of confounding biases, e.g., with regard to selection and ascertainment. Uncontrolled trials using progression-free survival or time to progression as a primary end point should be considered on a case-by-case basis, and the methodology to be applied should be thoroughly described in the protocol.

4. Response Review

For trials where the response rate is the primary end point, it is strongly recommended that all responses be reviewed by an expert or experts independent of the study at the study's completion. Simultaneous review of the patients' files and radiologic images is the best approach.

(*Note:* When a review of the radiologic images is to take place, it is also recommended that images be free of marks that might obscure the lesions or bias the evaluation of the reviewer[s]).

5. Reporting of Results

All patients included in the study must be assessed for response to treatment, even if there are major protocol treatment deviations or if they are ineligible. Each patient will be assigned one of the following categories: 1) complete response, 2) partial response, 3) stable disease, 4) progressive disease, 5) early death from malignant disease, 6) early death from toxicity, 7) early death because of other cause, or 9) unknown (not assessable, insufficient data). (*Note:* By arbitrary convention, category 9 usually designates the "unknown" status of any type of data in a clinical database.)

All of the patients who met the eligibility criteria should be included in the main analysis of the response rate. Patients in response categories 4–9 should be considered as failing to respond to treatment (disease progression). Thus, an incorrect treatment schedule or drug administration does not result in exclusion from the analysis of the response rate. Precise definitions for categories 4–9 will be protocol specific.

All conclusions should be based on all eligible patients.

Subanalyses may then be performed on the basis of a subset of patients, excluding those for whom major protocol deviations have been identified (e.g., early death due to other reasons, early discontinuation of treatment, major protocol violations, etc). However, these subanalyses may not serve as the basis for drawing conclusions concerning treatment efficacy, and the reasons for excluding patients from the analysis should be clearly reported. The 95% confidence intervals should be provided.

6. Response Evaluation in Randomized Phase III Trials

Response evaluation in phase III trials may be an indicator of the relative antitumor activity of the treatments evaluated but may usually not solely predict the real therapeutic benefit for the population studied. If objective response is selected as a primary end point for a phase III study (only in circumstances where a direct relationship between objective tumor response and a real therapeutic benefit can be unambiguously demonstrated for the population studied), the same criteria as those applicable to phase II trials (RECIST guidelines) should be used.

On the other hand, some of the guidelines presented in this special article might not be required in trials, such as phase III trials, in which objective response is *not* the primary end point. For example, in such trials, it might not be necessary to measure as many as 10 target lesions or to confirm response with a follow-up assessment after 4 weeks or more. Protocols should be written clearly with respect to planned response evaluation and whether confirmation is required so as to avoid *post-hoc* decisions affecting patient evaluability.

Appendix I. Specifications for Radiologic Imaging

These notes are recommendations for use in clinical studies and, as such, these protocols for computed tomography (CT) and magnetic resonance imaging (MRI) scanning may differ from those employed in clinical practice at various institutions. The use of standardized protocols allows comparability both within and between different studies, irrespective of where the examination has been undertaken.

Specific Notes

- For chest x-ray, not only should the film be performed in full inspiration in the posteroanterior projection, but also the film to tube distance should remain constant between examinations. However, patients in trials with advanced disease may not be well enough to fulfill these criteria, and such situations should be reported together with the measurements.

Lesions bordering the thoracic wall are not suitable for measurements by chest x-ray, since a slight change in position of the patients can cause considerable differences in the plane in which the lesion is projected and may appear to cause a change that is actually an artifact. These lesions should be followed by a CT or an MRI. Similarly, lesions bordering or involving the mediastinum should be documented on CT or MRI.

- CT scans of the thorax, abdomen, and pelvis should be contiguous throughout the anatomic region of interest. As a rule of thumb, the minimum size of the lesion should be no less than double the slice thickness. Lesions smaller than this are subject to substantial "partial volume" effects (i.e., size is underestimated because of the distance of the cut from the longest diameter; such a lesion may appear to have responded or progressed on subsequent examinations, when, in fact, they remain the same size [Fig. 1]). This minimum lesion size for a given slice thickness at baseline ensures that any lesion appearing smaller on subsequent examinations will truly be decreasing in size. The longest diameter of each target lesion should be selected in the axial plane only.

The type of CT scanner is important regarding the slice thickness and minimum-sized lesion. For spiral (helical) CT scanners, the minimum size of any given lesion at baseline may be 10 mm, provided the images are reconstructed contiguously at 5-mm intervals. For conventional CT scanners, the minimum-sized lesion should be 20 mm by use of a contiguous slice thickness of 10 mm.

The fundamental difference between spiral and conventional CT is that conventional CT acquires the information only for the particular slice thickness scanned, which is then expressed as a two-dimensional representation of that thickness or volume as a gray scale image. The next slice thickness needs to be scanned before it can be imaged and so on. Spiral CT acquires the data for the whole volume imaged, typically the whole of the thorax or upper abdomen in a single breath hold of about 20–30 seconds. To view the images, a suitable reconstruction algorithm is selected, by the machine, so the data are appropriately imaged. As suggested above, for spiral CT, 5-mm reconstructions can be made, thereby allowing a minimum-sized lesion of 10 mm.

Spiral CT is now the standard in most hospitals involved in cancer

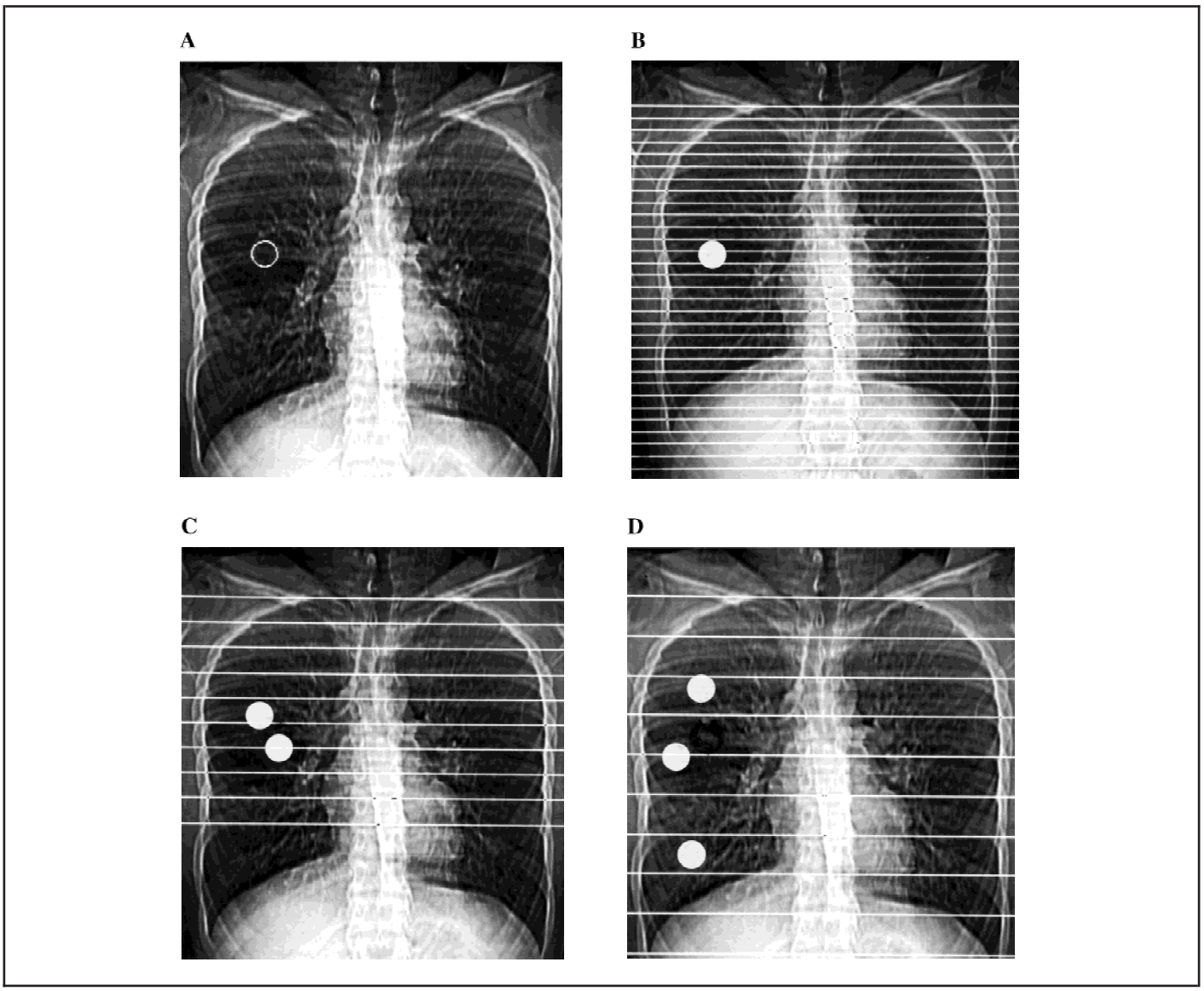


Fig 1. A) Computed tomography (CT) "scannogram" of the thorax with a simulated 20-mm lesion in the right mid-zone. **B**) CT "scannogram" of the thorax with contiguous slices of 10-mm thickness. Each volume within the slice thickness is scanned, and the average attenuation coefficient (i.e., density of multiple small cubes [voxels]) is represented spatially in two dimensions (pixels) as a cross-sectional image on a gray scale. It is important to note each line on the figure is a spatial representation of the average density for the structures that pass through that slice thickness, and the line does not represent a thin "cut" through it at that level. Therefore, a lesion of at least 20 mm will appear about its true diameter on at least one image because sufficient volume of the lesion is present so as not to average it down substantially. **C**) CT scannogram performed at 15-mm intervals. Depending on how much of the tumor is within the slice thickness, the average density may be substantially underestimated, as in the upper of the two lesions, or it may approximate the true tumor diameter, lower lesion. This is an oversimplification of the process but illustrates the point without going into the physics of CT reconstruction. **D**) CT scannogram performed at 24-mm intervals and of 10-mm thickness. The lesion may be imaged through its diameter, it may be partially imaged, or it may not be imaged at all. This is the equivalent of imaging a very small lesion and trying to determine whether its true diameter has changed from one examination to the next.

management in the United States, Europe, and Japan, so the above comments related to spiral CT are pertinent. However, some institutions involved in clinical trials will have conventional CT, but the number of these scanners will decline as they are replaced by spiral CT.

Other body parts, where CT scans are of different slice thickness (such as the neck, which is typically 5-mm thickness), or in the young pediatric population, where the slice thickness may be different, the minimum-sized lesion allowable for measurability of the lesion may be different. However, it should be double the slice thickness. The slice thickness and the minimum-sized lesion should be specified in the study protocol.

In patients in whom the abdomen and pelvis have been imaged, oral contrast agents should be given to accentuate the bowel against other soft-tissue masses. This procedure is almost universally undertaken on a routine basis.

Intravenous contrast agents should also be given, unless contraindicated for medical reasons such as allergy. This is to accentuate vascular structures from adjacent lymph node masses and to help enhance liver and other visceral metastases. Although, in clinical practice, its use may add little, in the context of a clinical study where objective response rate based on measurable disease is the end point, unless an intravenous contrast agent is given, a substantial number of otherwise measurable lesions will not be measurable. The use of intravenous contrast agents may sometimes seem unnecessary to monitor the evolution of specific disease sites (e.g., in patients in whom the disease is apparently restricted to the periphery of the lungs). However, the aim of a clinical

study is to ensure that lesions are truly resolving, and there is no evidence of new disease at other sites scanned (e.g., small metastases in the liver) that may be more easily demonstrated with the use of intravenous contrast agent that should, therefore, also be considered in this context.

The method of administration of intravenous contrast agents is variable. Rather than try to institute rigid rules regarding methods for administering contrast agents and the volume injected, it is appropriate to suggest that an adequate volume of a suitable contrast agent should be given so that the metastases are demonstrated to best effect and a consistent method is used on subsequent examinations for any given patient.

All images from each examination should be included and not "selected" images of the apparent lesion. This distinction is intended to ensure that, if a review is undertaken, the reviewer can satisfy himself/herself that no other abnormalities coexist. All window settings should be included, particularly in the thorax, where the lung and soft-tissue windows should be considered.

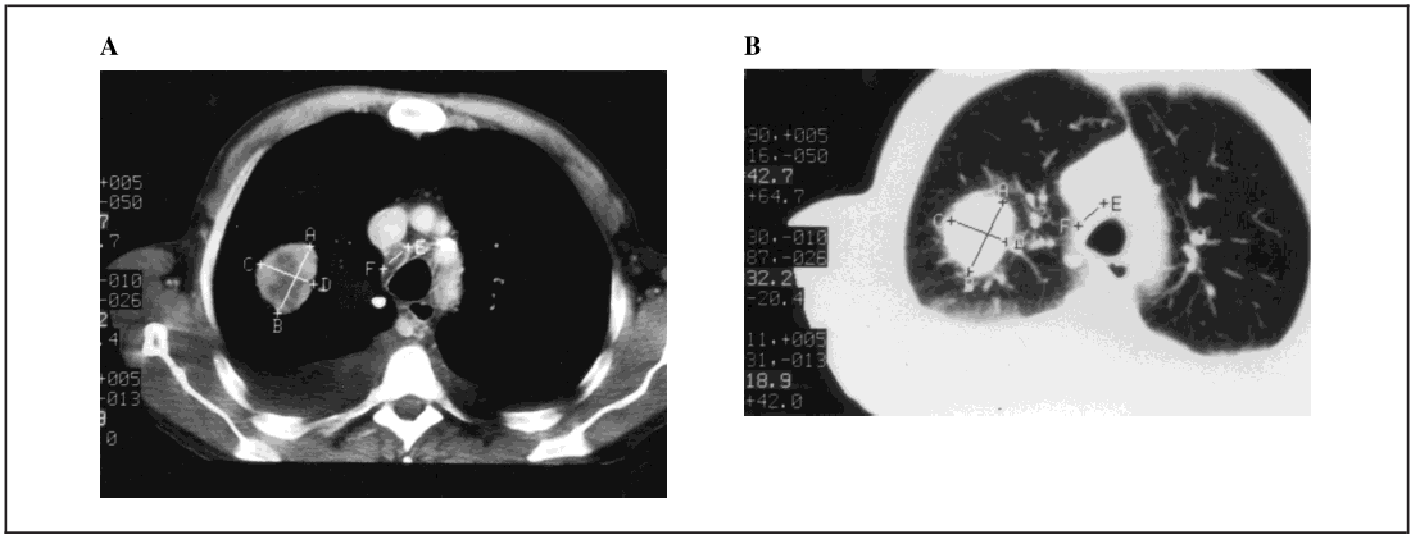


Fig 2. A) Computed tomography (CT) scan of the thorax at the level of the carina on "soft-tissue" windows. Two lesions have been measured with calipers. The intraparenchymal lesion has been measured bidimensionally, using the greatest diameter and the greatest perpendicular distance. Unidimensional measurements require only the greatest diameter to be measured. The anterior–carinal lymph node has been measured using unidimensional criteria. **B**) The same image as above imaged on "lung" windows, with the calipers remaining as they were for the soft-tissue measurements. The size of the lung lesion appears different. The anterior–carinal lymph node cannot be measured on these windows. The same windows should be used on subsequent examinations to measure any lesions. Some favor soft-tissue windows, so paratracheal, anterior, and subcarinal lesions may be followed on the same settings as intraparenchymal lesions.

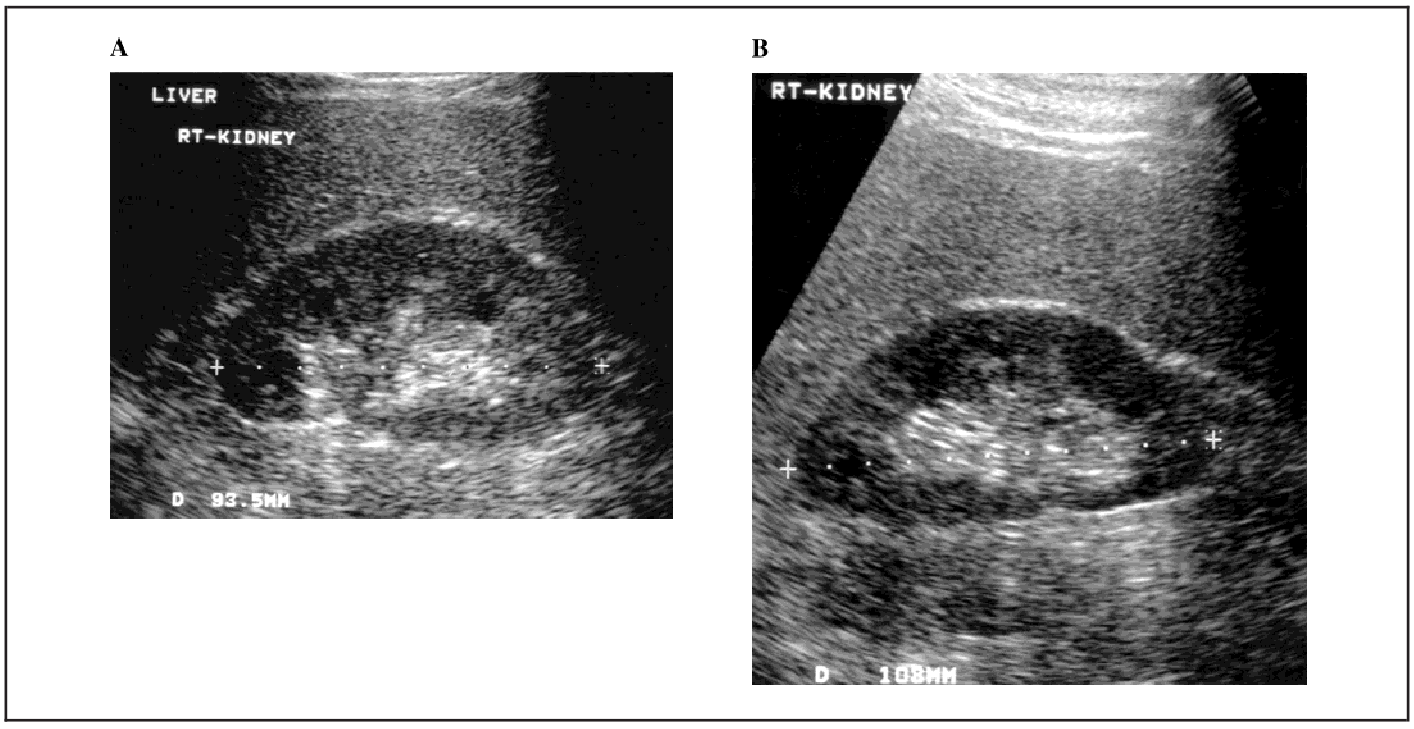


Fig 3. A) Ultrasound scan of a normal structure, the right kidney, which has been measured as 93 mm with the use of callipers. **B**) Ultrasound scan of the same kidney taken a few minutes later when it measures 108 mm. It appears to have increased in size by 16%. The difference is due to foreshortening of the kidney in panel A. The lack of anatomic landmarks makes accurate measurement in the same plane on subsequent examinations difficult. One has to hope that the measurements given on the hard copy film are a true and accurate reflection of events.

Lesions should be measured on the same window setting on each examination. It is not acceptable to measure a lesion on lung windows on one examination and on soft-tissue settings on the next (Fig. 2). In the lung, it does not really matter whether lung or soft-tissue windows are used for intraparenchymal lesions, provided a thorough assessment of nodal and parenchymal disease has been undertaken and the target lesions are measured as appropriate by use of the same window settings for repeated examinations throughout the study.

- Use of MRI is a complex issue. MRI is entirely acceptable and capable of providing images in different anatomic planes. It is, therefore, important that, when MRI is used, lesions must be measured in the same anatomic plane by use of the same imaging sequences on subsequent examinations. MRI scanners vary in the images produced. Some of the factors involved include the magnet strength (high-field magnets require shorter scan times, typically 2–5 minutes), the coil design, and patient cooperation. Wherever possible, the same scanner should be used. For instance, the images provided by a 1.5-Tesla scanner will differ from those provided by a 0.5-Tesla scanner. Although comparisons can be made between images from different scanners, such comparisons are not ideal. Moreover, many patients with advanced malignancy are in pain, so their ability to remain still for the duration of a scan sequence—on the order of 2–5 minutes—is limited. Any movement during the scan time leads to motion artifacts and degradation of image quality, so that the examination will probably be useless. For these reasons, CT is, at this point in time, the imaging modality of choice.
- Ultrasound examinations should not be used in clinical trials to measure tumor regression or progression of lesions that are not superficial because the examination is necessarily subjective. Entire examinations cannot be reproduced for independent review at a later date, and it must be assumed, whether or not it is the case, that the hard-copy films available represent a true and accurate reflection of events (Fig. 3). Furthermore, if, for example, the only measurable lesion is in the para-aortic region of the abdomen and if gas in the bowel overlies the lesion, the lesion will not be detected because the ultrasound beam cannot penetrate the gas. Accordingly, the disease staging (or restaging for treatment evaluation) for this patient will not be accurate.

The same imaging modality must be used throughout the study to measure disease. Different imaging techniques have differing sensitivities, so any given lesion may have different dimensions at any given time if measured with different modalities. It is, therefore, not acceptable to interchange different modalities throughout a trial and use these measurements. It must be the same technique throughout.

It is desirable to try to standardize the imaging modalities without adding undue constraints so that patients are not unnecessarily excluded from clinical trials.

Appendix II. Relationship Between Change in Diameter, Product, and Volume

Appendix II, Table 2. Relationship between change in diameter, product, and volume*

	Diameter, $2r$	Product, $(2r)^2$	Volume, $4/3\pi r^3$
Response	Decrease	Decrease	Decrease
	30%	50%	65%
	50%	75%	87%
Disease progression	Increase	Increase	Increase
	12%	25%	40%
	20%	44%	73%
	25%	56%	95%
	30%	69%	120%

*Shaded areas represent the response evaluation criteria in solid tumors (diameter) and World Health Organization (product) criteria for change in tumor size to meet response and disease progression definitions.

Appendix III. Response Evaluation Criteria in Solid Tumors (RECIST) Working Group and Special Acknowledgments

RECIST Working Group

P. Therasse (Chair), J. Verweij, M. Van Glabbeke, A. T. van Oosterom, European Organization for Research and Treatment of Cancer (Brussels, Belgium); S. G. Arbuck, R. S. Kaplan, M. C. Christian, National Cancer Institute, United States (Bethesda, MD); E. Eisenhauer, National Cancer Institute of Canada Clinical Trials Group (Kingston); S. Gwyther, East Surrey Hospital (Redhill, U.K.); and J. Wanders, New Drug Development Office Oncology (Amsterdam, The Netherlands).

Retrospective Analyses

L. A. Rubinstein, National Cancer Institute, United States; B. K. James, A. Muldal, W. Walsh, National Cancer Institute of Canada Clinical Trials Group; S. Green, Southwest Oncology Group (Seattle, WA); M. Terenziani, National Cancer Institute (Milan, Italy); D. Vena, Emmes Corporation (Rockville, MD); R. Canetta, J. Burroughs, Bristol-Myers Squibb (Wallingford, CT); A. Riva, M. Murawsky, Rhone-Poulenc Rorer Pharmaceuticals Inc. (Paris, France).

Appendix IV. Participants in the October 1998 Workshop to Develop the Final Response Evaluation Criteria in Solid Tumors (RECIST) Document and Further Acknowledgments

Participants

S. C. S. Kao, Children's Cancer Study Group (Iowa City, IA); D. Grinblatt, Cancer and Leukemia Group B (CALGB) (Chicago, IL); B. Giantonio, Eastern Cooperative Oncology Group (ECOG) (Philadelphia, PA); F. B. Stehman, Gynecologic Oncology Group (GOG) (Indianapolis, IN); A. Trotti, Radiation Therapy Oncology Group (Tampa, FL); C. A. Coltman, Southwest Oncology Group (SWOG) (San Antonio, TX); R. E. Smith, National Surgical Adjuvant Breast and Bowel Project (Pittsburgh, PA); J. Zalcberg, Peter MacCallum Cancer Institute (Melbourne), Australia; N. Saijo, National Cancer Center Hospital (Tokyo, Japan); Y. Fujiwara, National Institute of Health Sciences (Tokyo); G. Schwartsmann, Hospital de Clinicas de Porto Alegre (Brazil); A. Klein, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa, ON); B. Weinerman, National Cancer Institute of Canada Clinical Trials Group (Kingston, ON); D. Warr, Ontario Cancer Institute/Princess Margaret Hospital (Toronto); P. Liati, South Europe New Drugs Organization (Milan, Italy); S. Einstein, Bio-Imaging Technologies (West Trenton, NJ); S. Négrier, L. Ollivier, Fédération Nationale des Centres de Lutte contre le Cancer (Paris, France); M. Marty, International Cancer Cooperative Group/French Drug Agency (Paris); H. Anderson, A. R. Hanauske, European Organization for Research and Treatment of Cancer (EORTC) (Brussels, Belgium); M. R. Mirza, Odense University Hospital (Denmark); J. Ersboll, The European Agency for the Evaluation of Medicinal Products (Bronshoj, Denmark); C. Pagonis, Cancer Research Campaign (London, U.K.); S. Hatty, Eli Lilly and Co., (Surrey, U.K.); A. Riva, Rhone-Poulenc Rorer Pharmaceuticals Inc. (Paris); C. Royce, GlaxoWellcome (Middlesex, U.K.); G. Burke, Novartis Pharma AG (Basel, Switzerland); I. Horak, Janssen Research Foundation (Beerse, Belgium); G. Hoctin-Boes, Zeneca (Macclesfield Cheshire, U.K.); C. Weil, Bristol-Myers Squibb (Waterloo, Belgium); M. G. Zurlo, Pharmacia & Upjohn (Milan); S. Z. Fields, SmithKline Beecham Pharmaceuticals (Collegeville, PA); B. Osterwalder, Hoffmann-La Roche Inc. (Basel); Y. Shimamura, Taiho Pharmaceutical Co. Ltd. (Tokyo); and M. Okabe, Kyowa-Hakko-Kogyo Co. Ltd. (Tokyo).

Additional comments were received from the following:

A. Hamilton, R. De Wit, E. Van Cutsem, J. Wils, J.-L. Lefèbvre, I. Vergote, M. S. Aapro, J.-F. Bosset, M. Hernandez-Bronchud, D. Lacombe, H. J. Schmoll, E. Van Limbergen, P. Fumoleau, A. Bowman, U. Bruntsch, EORTC (Brussels); B. Escudier, P. Thiesse, N. Tournemaine, P. Troufleau, C. Lasset, F. Gomez, Fédération Nationale des Centres de Lutte contre le Cancer (Paris); G. Rustin, Mount Vernon Hospital (Northwood Middlesex, U.K.); S. B. Kaye, Western Infirmary (Glasgow, U.K.); A. Goldhirsch, F. Nolè, G. Zampino, F. De Braud, M. Colleoni, E. Munzone, T. De Pas, International Breast Cancer Study Group and Istituto Europeo di Oncologia (Milan); M. Castiglione, J. F. Delaloye, A. Roth, C. Sessa, D. Hess, B. Thürlimann, C. Böhme, T. Cerny, U. Hess, Schweizer Arbeitsgemeinschaft für Klinische Krebsforschung (Bern, Switzerland); H. J. Stewart, Scottish Cancer Therapy Network (Edinburgh, U.K.); A. Howell, J. F. R. Robertson, United Kingdom Coordinating Committee on Cancer Research (Nottingham); K. Noever, Bio-Imaging Technologies (Monheim, Germany); M. Kurihara, Toyosu Hospital, SHOWA University (Tokyo); L. Seymour, J. Pater, J. Rusthoven, F. Shepherd, J. Maroun, G. Cairncross, D. Stewart, K. Pritchard, National Cancer Institute of Canada Clinical Trials Group (Kingston); T. Uscinowicz, Health Canada, Bureau of Pharmaceutical Assessment (Ottawa); I. Tannock, Princess Margaret Hospital (Toronto); M. Azab, QLT Phototherapeutics (Vancouver, Canada); V. H. C. Bramwell, Canadian Sarcoma Group (London); P. O'Dwyer, ECOG (Philadelphia); A. Martin, S. Ellenberg, U.S. Food and Drug Administration (Rockville, MD); C. Chow, D. Sullivan, A. Murgo, A. Dwyer, J. Tatum, National Cancer Institute (Bethesda, MD); R. Schilsky, CALGB (Chicago, IL); J. Crowley, S. Green, SWOG (Seattle, WA); R. Park, GOG (Philadelphia, PA); V. Land, B. D. Fletcher, Pediatric Oncology Group (Chicago, IL); B. Hillman, University of Virginia (Charlottesville); F. Muggia, New York University Medical Center (New York); C. Erlichman, Mayo Clinic (Rochester, MN); L. H. Schwartz, Memorial Sloan-Kettering Cancer Center (New York, NY); S. P. Balcerzak, Ohio State University Health Sciences Center (Columbus); G. Fleming, CALGB (Chicago); G. Sorensen, Harvard University (Cambridge, MA); H. Levy, Thomas Jefferson University (Philadelphia); N. Patz, Duke University (Durham, NC); C. Visseren-Grul, Eli Lilly Nederland BV (Nieuwegein, The Netherlands)/J. Walling, Lilly Research Laboratories (Indianapolis); P. Hellemans, Janssen Research Foundation (Beerse, Belgium); L. Finke, Merck (Darmstadt, Germany); A. Man, N. Barbet, Novartis Pharma AG (Basel); G. Massimini, Pharmacia & Upjohn (Milan); J, Jimeno, Pharma Mar (Madrid, Spain); I. Hudson, SmithKline Beecham Pharmaceuticals (Essex, U.K.); and J. Krebs, R. A. Beckman, S. Lane, D. Fitts, SmithKline Beecham Pharmaceuticals (Collegeville).

APPENDIX V. RETROSPECTIVE COMPARISON OF RESPONSE/DISEASE PROGRESSION RATES OBTAINED WITH THE WORLD HEALTH ORGANIZATION (WHO)/SOUTHWEST ONCOLOGY GROUP CRITERIA AND THE NEW RESPONSE EVALUATION CRITERIA IN SOLID TUMORS (RECIST) CRITERIA

To evaluate the hypothesis by which unidimensional measurement of tumor lesions may substitute for the usual bidimensional approach, a number of retrospective analyses have been undertaken. The results of these analysis are given below in this section.

1. Comparison of Response and Disease Progression Rates by Use of WHO (or Modified WHO) or RECIST Methods

1.1. Trials Evaluated

No specific selection criteria were employed except that trial data had to include serial (repeated) records of tumor measurements. Several groups evaluated their own data on one or more such studies (National Institute of Canada Clinical Trials Group, Kingston, ON; U.S. National Cancer Institute, Bethesda, MD; and Rhone-Poulenc Rorer Pharmaceuticals Inc., Paris, France) or made data available for evaluation to the U.S. National Cancer Institute (Southwest Oncology Group and Bristol-Myers Squibb, Wallingford, CT)

1.2. Response Criteria Evaluated

Not all databases were assessed for all response outcomes. At the outset of this process, the most interest was in the assessment of complete plus partial response rate comparisons by both the WHO and new RECIST criteria. Once these data suggested no impact of using the new criteria on the response rate, several more databases were analyzed for the impact of the use of the new criteria not only on complete response plus partial response but also on stable disease and progressive disease rates (*see* Appendix V, Table 4) and on time to disease progression (*see* Appendix V, Table 5).

1.3. Methods of Comparison

For each patient in each study, baseline sums were calculated (sum of products of the two longest diameters in perpendicular dimensions for WHO and sum of longest diameters for RECIST). After each assessment, when new tumor measures were available, the sums were recalculated. Patients were assigned complete response, partial response, stable disease, and progressive disease as their "best" response on the basis of achieving the measurement criteria as indicated in Appendix V, Table 3. For both WHO and RECIST, a minimum interval of 4 weeks was required to consider complete response and partial response confirmed. Each patient could, therefore, be assigned a best response according to each of the two criteria. The overall response and disease progression rates could be calculated for the population studied for each trial or dataset examined.

(*Note:* For WHO progressive disease, as is the convention in most groups, an increase in sums of products was required, not an increase in only one lesion.)

1.4. Results

2. Evaluation of Time to Disease Progression

Time to disease progression was evaluated, comparing WHO criteria with RECIST in a dataset provided by the Southwest Oncology Group

Appendix V, Table 3. Definition of best response according to WHO or RECIST criteria*

Best response	WHO change in sum of products	RECIST change in sums longest diameters
CR	Disappearance; confirmed at 4 wks†	Disappearance; confirmed at 4 wks†
PR	50% decrease; confirmed at 4 wks†	30% decrease; confirmed at 4 wks†
SD	Neither PR nor PD criteria met	Neither PR nor PD criteria met
PD	25% increase; no CR, PR, or SD documented before increased disease	20% increase; no CR, PR, or SD documented before increased disease

*WHO = World Health Organization; RECIST = Response Evaluation Criteria in Solid Tumors; CR = complete response, PR = partial response, SD = stable disease, and PD = progressive disease.

†For the Bristol-Myers Squibb (Wallingford, CT) dataset, only unconfirmed CR and PR have been used to compare best response measured in one dimension (RECIST criteria) versus best response measured in two dimensions (WHO criteria). The computer flag identifying confirmed response in this dataset could not be used in the comparison for technical reasons.

Appendix V, Table 4. Comparison of RECIST (unidimensional) and WHO (bidimensional) criteria in the same patients recruited in 14 different trials*

Tumor site/type	Criteria	No. of patients evaluated	Best response				RR	PD rate
			CR	PR	SD	PD		
Breast†	WHO	48	4	22			54%	
	RECIST	48	4	22			54%	
Breast‡	WHO	172	4	36			23%	
	RECIST	172	4	40			26%	
Brain†	WHO	31	12	10			71%	
	RECIST	31	12	10			71%	
Melanoma†	WHO	190	9	37			24%	
	RECIST	190	9	34			23%	
Breast§	WHO	531	50	102			29%	
	RECIST	531	50	108			30%	
Colon§	WHO	1096	12	137			14%	
	RECIST	1096	12	133			13%	
Lung§	WHO	1197	60	317			32%	
	RECIST	1197	60	318			32%	
Ovary§	WHO	554	24	108			24%	
	RECIST	554	24	105			23%	
Lung†	WHO	24	0	4	16	4	17%	17%
	RECIST	24	0	4	19	1	17%	4%
Colon†	WHO	31	1	6	15	9	23%	29%
	RECIST	31	1	5	16	9	21%	29%
Sarcoma†	WHO	28	1	4	13	10	18%	36%
	RECIST	28	1	5	17	5	21%	18%
Ovary†	WHO	45	0	7	19	19	16%	42%
	RECIST	45	0	6	21	18	13%	40%
Breast‖	WHO	306	18	114	117	57	43%	19%
	RECIST	306	18	108	124	56	41%	18%
Breast‖	WHO	360	10	73	135	142	23%	39%
	RECIST	361	10	70	139	142	22%	39%
Total (all studies where tumor response was evaluated)	WHO	4613	205	977			25.6%	
	RECIST	4614	205	968			25.4%	
Total (all studies where PD as well as CR + PR were evaluated)	WHO	794			315	241		30.3%
	RECIST	795			336	231		29%

*WHO = World Health Organization *(3);* RECIST = Response Evaluation Criteria in Solid Tumors; CR = complete response; PR = partial response; SD = stable disease; PD = progressive disease; and RR = response rate.

†Data from the National Cancer Institute of Canada Clinical Trials Group phase II and III trials.

‡Data from the National Cancer Institute, United States phase III trial.

§Data from Bristol-Myers Squibb (Wallingford, CT) phase II and III trials.

‖Data from Rhone-Poulenc Rorer Pharmaceuticals Inc., (Paris, France) phase III trials (*note:* one patient in this database had unidimensional measured lesions only and could not be evaluated with the WHO criteria).

Appendix V, Table 5. Proportions of patients with disease progression by different assessment methods*

	No. of patients	%
Total No. of progressors	234	100
Progress by appearance of new lesions†	118	50
Progress by increase in pre-existing measurable disease	116	50
Same date of disease progression by WHO and RECIST criteria	215	91.9
Different date of disease progression	19	8.1
Earlier PD with WHO criterion	17	7.3
Earlier PD with unidimensional criterion	2	0.9

*PD = progressive disease; WHO = World Health Organizaiton; and RECIST = Response Evaluation Criteria in Solid Tumors.

†Also includes a few patients with PD because of marked increase of nonmeasurable disease.

Appendix V, Table 6. Magnitude of time to disease progression disagreements when differences existed*

	No. of patients	% (of 234, *see* above)
No. of progressors with differing progression dates	19	8.1
8–9 wks' difference	3	1.3
12 wks' difference	1	0.4
24–31 wks' difference†	2	0.9
Difference uncertain due to censoring of either WHO or RECIST progression time‡	13	5.6

*WHO = World Health Organization; RECIST = Response Evaluation Criteria in Solid Tumors.

†For one patient, progression by RECIST (one-dimension) criteria preceded that by WHO criteria by 24 weeks due primarily to one-dimensional growth. For a second patient, with a colon tumor that increased in cross-section by 25%, then regressed completely, and then recurred, progression by WHO criteria preceded that by RECIST criteria by 31 weeks.

‡As indicated in Appendix V, Table 6, 13 of the 19 patients had uncertain disease progression time differences when comparing RECIST and WHO criteria. In these patients, the RECIST progression criteria were not met by the time that disease progression by Southwest Oncology Group (SWOG) criteria *(5)* had occurred (50% increase or a 10 cm^2 increase in tumor cross-section). Notably, six of these patients had the same disease progression dates determined by use of WHO (25% bidimensional increase) and SWOG (50% bidimensional increase) criteria. Since 20% unidimensional increase (RECIST) is equivalent to approximately 44% bidimensional increase, it is likely, although not certain, that disease progression by RECIST unidimensional criteria would have occurred soon after disease progression by SWOG and WHO criteria. For three patients, the difference between the WHO and SWOG 50% bidimensional increase was 10–12 weeks. Again, it is likely, although it cannot be proven, that RECIST criteria would have been met soon after. The remaining four of the 13 patients where difference between WHO and RECIST progression times are uncertain were categorized as progressive disease following SWOG's criteria *(5)* because of an increase of the tumor surface of greater than or equal to 10 cm^2. For these patients, the magnitude of the difference is entirely uncertain.

(SWOG). Since SWOG criteria *(5)* for disease progression is a 50% increase in the sum of the products, or new disease, or an absolute increase of 10 cm^2 in the sum of the products, this dataset provided the means of assessing the impact of time to disease progression differences between a 25% increase in the sum of the products and a 20% increase in the sum of the longest diameters (equivalent to approximately a 44% increase in the product sum).

2.1. Dataset Evaluated

The dataset includes 234 patients with progressive disease as defined by the SWOG *(5)*. All patients had baseline measurable disease followed by the same technique(s) until disease progression. The tumor types included were melanoma and colorectal, lung, and breast cancers.

REFERENCES

(1) Zubrod CG, Schneiderman SM, Frei E III, Brindley C, Gold GL, Schnider B, et al. Appraisal of methods for the study of chemotherapy of cancer in man: comparative therapeutic trial of nitrogen mustard and thio phosphoamide. J Chronic Dis 1960;11:7–33.

(2) Gehan E, Schneidermann M. Historical and methodological developments in clinical trials at the National Cancer Institute. Stat Med 1990;9: 871–80.

(3) WHO handbook for reporting results of cancer treatment. Geneva (Switzerland): World Health Organization Offset Publication No. 48; 1979.

(4) Miller AB, Hogestraeten B, Staquet M, Winkler A. Reporting results of cancer treatment. Cancer 1981;47:207–14.

(5) Green S, Weiss GR. Southwest Oncology Group standard response criteria, endpoint definitions and toxicity criteria. Invest New Drugs 1992;10:239–53.

(6) James K, Eisenhauer E, Christian M, Terenziani M, Vena D, Mudal A, et al. Measuring response in solid tumors: unidimensional versus bidimensional measurement. J Natl Cancer Inst 1999;91:523–8.

NOTE

Manuscript received June 24, 1999; revised November 30, 1999; accepted December 3, 1999.

Response Evaluation Criteria in Solid Tumors (RECIST) Quick Reference:

Eligibility

- Only patients with measurable disease at baseline should be included in protocols where objective tumor response is the primary endpoint.

  Measurable disease - the presence of at least one measurable lesion. If the measurable disease is restricted to a solitary lesion, its neoplastic nature should be confirmed by cytology/histology**.**

  Measurable lesions - lesions that can be accurately measured in at least one dimension with longest diameter ≥20 mm using conventional techniques or ≥10 mm with spiral CT scan.

  Non-measurable lesions - all other lesions, including small lesions (longest diameter <20 mm with conventional techniques or <10 mm with spiral CT scan), i.e., bone lesions, leptomeningeal disease, ascites, pleural/pericardial effusion, inflammatory breast disease, lymphangitis cutis/pulmonis, cystic lesions, and also abdominal masses that are not confirmed and followed by imaging techniques; and**.**

- All measurements should be taken and recorded in metric notation, using a ruler or calipers. All baseline evaluations should be performed as closely as possible to the beginning of treatment and never more than 4 weeks before the beginning of the treatment.

- The same method of assessment and the same technique should be used to characterize each identified and reported lesion at baseline and during follow-up.

- Clinical lesions will only be considered measurable when they are superficial (e.g., skin nodules and palpable lymph nodes). For the case of skin lesions, documentation by color photography, including a ruler to estimate the size of the lesion, is recommended**.**

Methods of Measurement –

- CT and MRI are the best currently available and reproducible methods to measure target lesions selected for response assessment. Conventional CT and MRI should be performed with cuts of 10 mm or less in slice thickness contiguously. Spiral CT should be performed using a 5 mm contiguous reconstruction algorithm. This applies to tumors of the chest, abdomen and pelvis. Head and neck tumors and those of extremities usually require specific protocols.

- Lesions on chest X-ray are acceptable as measurable lesions when they are clearly defined and surrounded by aerated lung. However, CT is preferable.

- When the primary endpoint of the study is objective response evaluation, ultrasound (US) should not be used to measure tumor lesions. It is, however, a possible alternative to clinical measurements of superficial palpable lymph nodes, subcutaneous lesions and thyroid nodules. US might also be useful to confirm the complete disappearance of superficial lesions usually assessed by clinical examination.

- The utilization of endoscopy and laparoscopy for objective tumor evaluation has not yet been fully and widely validated. Their uses in this specific context require sophisticated equipment and a high level of expertise that may only be available in some centers. Therefore, the utilization of such techniques for objective tumor response should be restricted to validation purposes in specialized centers. However, such techniques can be useful in confirming complete pathological response when biopsies are obtained.

- Tumor markers alone cannot be used to assess response. If markers are initially above the upper normal limit, they must normalize for a patient to be considered in complete clinical response when all lesions have disappeared.

- Cytology and histology can be used to differentiate between PR and CR in rare cases (e.g., after treatment to differentiate between residual benign lesions and residual malignant lesions in tumor types such as germ cell tumors).

Baseline documentation of "Target" and "Non-Target" lesions

- All measurable lesions up to a maximum of five lesions per organ and 10 lesions in total, representative of all involved organs should be identified as ***target lesions*** and recorded and measured at baseline.

- Target lesions should be selected on the basis of their size (lesions with the longest diameter) and their suitability for accurate repeated measurements (either by imaging techniques or clinically).

- A sum of the longest diameter (LD) for *all target lesions* will be calculated and reported as the baseline sum LD. The baseline sum LD will be used as reference by which to characterize the objective tumor.

- All other lesions (or sites of disease) should be identified as ***non-target lesions*** and should also be recorded at baseline. Measurements of these lesions are not required, but the presence or absence of each should be noted throughout follow-up.

Response Criteria

Evaluation of target lesions

* Complete Response (CR):	Disappearance of all target lesions
* Partial Response (PR):	At least a 30% decrease in the sum of the LD of target lesions, taking as reference the baseline sum LD
* Progressive Disease (PD):	At least a 20% increase in the sum of the LD of target lesions, taking as reference the smallest sum LD recorded since the treatment started or the appearance of one or more new lesions
* Stable Disease (SD):	Neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD, taking as reference the smallest sum LD since the treatment started

Evaluation of non-target lesions

* Complete Response (CR):	Disappearance of all non-target lesions and normalization of tumor marker level
* Incomplete Response/ Stable Disease (SD):	Persistence of one or more non-target lesion(s) or/and maintenance of tumor marker level above the normal limits
* Progressive Disease (PD):	Appearance of one or more new lesions and/or unequivocal progression of existing non-target lesions (1)

(1) Although a clear progression of "non target" lesions only is exceptional, in such circumstances, the opinion of the treating physician should prevail and the progression status should be confirmed later on by the review panel (or study chair).

Evaluation of best overall response

The best overall response is the best response recorded from the start of the treatment until disease progression/recurrence (taking as reference for PD the smallest measurements recorded since the treatment started). In general, the patient's best response assignment will depend on the achievement of both measurement and confirmation criteria

Target lesions	Non-Target lesions	New Lesions	Overall response
CR	CR	No	CR
CR	Incomplete response/SD	No	PR
PR	Non-PD	No	PR
SD	Non-PD	No	SD
PD	Any	Yes or No	PD
Any	PD	Yes or No	PD
Any	Any	Yes	PD

- Patients with a global deterioration of health status requiring discontinuation of treatment without objective evidence of disease progression at that time should be classified as having “symptomatic deterioration”. Every effort should be made to document the objective progression even after discontinuation of treatment.

- In some circumstances it may be difficult to distinguish residual disease from normal tissue. When the evaluation of complete response depends on this determination, it is recommended that the residual lesion be investigated (fine needle aspirate/biopsy) to confirm the complete response status.

Confirmation

- The main goal of confirmation of objective response is to avoid overestimating the response rate observed. In cases where confirmation of response is not feasible, it should be made clear when reporting the outcome of such studies that the responses are not confirmed.

- To be assigned a status of PR or CR, changes in tumor measurements must be confirmed by repeat assessments that should be performed no less than 4 weeks after the criteria for response are first met. Longer intervals as determined by the study protocol may also be appropriate.

- In the case of SD, follow-up measurements must have met the SD criteria at least once after study entry at a minimum interval (in general, not less than 6-8 weeks) that is defined in the study protocol

Duration of overall response

- The duration of overall response is measured from the time measurement criteria are met for CR or PR (whichever status is recorded first) until the first date that recurrence or PD is objectively documented, taking as reference for PD the smallest measurements recorded since the treatment started.

Duration of stable disease

- SD is measured from the start of the treatment until the criteria for disease progression are met, taking as reference the smallest measurements recorded since the treatment started.

- The clinical relevance of the duration of SD varies for different tumor types and grades. Therefore, it is highly recommended that the protocol specify the minimal time interval required between two measurements for determination of SD. This time interval should take into account the expected clinical benefit that such a status may bring to the population under study.

Response review

- For trials where the response rate is the primary endpoint it is strongly recommended that all responses be reviewed by an expert(s) independent of the study at the study's completion. Simultaneous review of the patients' files and radiological images is the best approach.

Reporting of results

- All patients included in the study must be assessed for response to treatment, even if there are major protocol treatment deviations or if they are ineligible. Each patient will be assigned one of the following categories: 1) complete response, 2) partial response, 3) stable disease, 4) progressive disease, 5) early death from malignant disease, 6) early death from toxicity, 7) early death because of other cause, or 9) unknown (not assessable, insufficient data).

- All of the patients who met the eligibility criteria should be included in the main analysis of the response rate. Patients in response categories 4-9 should be considered as failing to respond to treatment (disease progression). Thus, an incorrect treatment schedule or drug administration does not result in exclusion from the analysis of the response rate. Precise definitions for categories 4-9 will be protocol specific.

- All conclusions should be based on all eligible patients.

- Subanalyses may then be performed on the basis of a subset of patients, excluding those for whom major protocol deviations have been identified (e.g., early death due to other reasons, early discontinuation of treatment, major protocol violations, etc.). However, these subanalyses may not serve as the basis for drawing conclusions concerning treatment efficacy, and the reasons for excluding patients from the analysis should be clearly reported.

- The 95% confidence intervals should be provided.