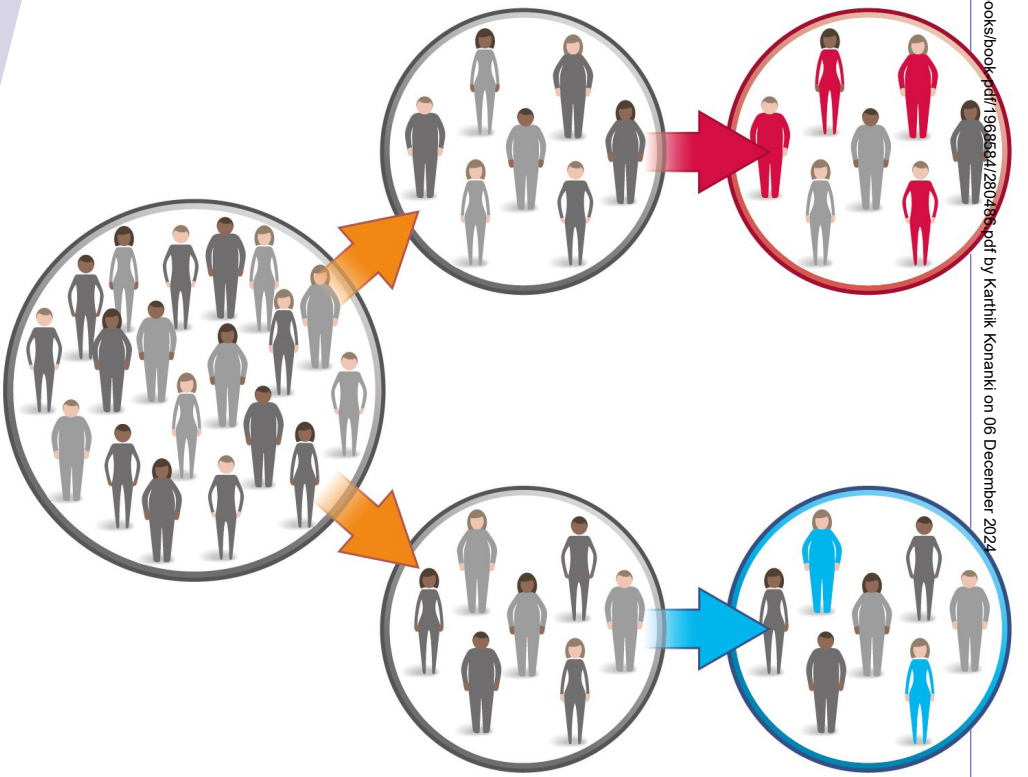
FAST FACTS

Clinical Trials in Oncology

Allan Hackshaw and Gavin CE Stuart



The fundamentals of design, conduct and interpretation

FAST FACTS

Clinical Trials in Oncology

Allan Hackshaw MSc
Professor of Epidemiology and Medical Statistics,
and Deputy Director, Cancer Research UK
& UCL Cancer Trials Centre
University College London
London, UK

Gavin CE Stuart MD FRCSC
Professor of Obstetrics and Gynaecology
University of British Columbia
Department of Gynaecology
Diamond Health Care Centre
Vancouver, Canada

Declaration of Independence
This book is as balanced and as practical as we can make it.
Ideas for improvement are always welcome: fastfacts@karger.com

KARGER

Fast Facts: Clinical Trials in Oncology
First published 2021

S. Karger Publishers Ltd, Elizabeth House, Queen Street, Abingdon,
Oxford OX14 3LN, UK
Tel: +44 (0)1235 523233

Book orders can be placed by telephone or email, or via the website.
Please telephone +41 61 306 1440 or email orders@karger.com
To order via the website, please go to karger.com

Fast Facts is a trademark of S. Karger Publishers Ltd.

A CIP record for this title is available from the British Library.

ISBN 978-1-912-77673-3

Hackshaw A (Allan)
Fast Facts: Clinical Trials in Oncology/
Allan Hackshaw, Gavin CE Stuart

Typesetting by Karger UK, Abingdon, UK.
Printed in the UK with Xpedient Print.

List of abbreviations

AE: adverse event

AR: adverse (drug) reaction

BICR: blinded independent central review

BSC: best supportive care

CI: confidence interval

CONSORT: Consolidated Standards of Reporting Trials

CR: complete response

CRF: case report form

CRM: continual reassessment method

CRO: contract research organization

CT: computed tomography

CTA: Clinical Trial Authorisation

CTCAE: Common Terminology Criteria for Adverse Events

DCB: duration of clinical benefit

DFS: disease-free survival

DLT: dose-limiting toxicity

DOR: duration of response

DSMB: Data and Safety Monitoring Board

ECOG: Eastern Cooperative Oncology Group

EFS: event-free survival

EMA: European Medicines Agency

EudraCT: EU Drug Regulating Authorities Clinical Trials (database)

FDA: US Food and Drug Administration

GCLP: Good Clinical Laboratory Practice

GCP: Good Clinical Practice

GDPR: General Data Protection Regulation

GMP: Good Manufacturing Practice

HR: hazard ratio

HRQoL: health-related quality of life

HTA: health technology assessment

IB: Investigator's Brochure

ICER: incremental cost-effectiveness ratio

ICH: International Conference on Harmonization

IDMC: Independent Data Monitoring Committee

IMP: Investigational Medicinal Product

IMPD: Investigational Medicinal Product Dossier

IMRT: intensity-modulated radiation therapy

IND: Investigational New Drug

IPD: individual patient data meta-analysis

IRB: Institutional Review Board

ITC: indirect treatment comparison

ITT: intention to treat

MED: minimum effective dose

MRI: magnetic resonance imaging

MTD: maximum tolerated dose

NGS: next-generation sequencing

ORR: overall (tumor) response rate

OS: overall survival

PD: pharmacodynamics

PD-L1: programmed death-ligand 1

PET: positron electron tomography

PFS: progression-free survival

PIS: Patient Information Sheet

PK: pharmacokinetics

PR: partial response

PRO: patient-reported outcome

PS: performance status

QALY: quality-adjusted life-year

QP: Qualified Person

REC: Research Ethics Committee

RECIST: Response Evaluation Criteria in Solid Tumors

RFS: relapse-free survival

RR: relative risk

RT: radiation therapy

RWD: real-world data

RWE: real-world evidence

SAE: serious adverse event

SAP: statistical analysis plan

SAR: serious adverse reaction

SD: stable disease

SOC: standard of care

SOP: standard operating procedure

SPC: Summary of Product Characteristics (sometimes abbreviated to SmPC)

SUSAR: suspected unexpected serious adverse reaction

TMF: Trial Master File

TTF: time to treatment failure

TTNT: time to next (anticancer) treatment

USM: urgent safety measure

Acknowledgment

The authors are grateful to Dr Claire Roddie, University College London, UK, for her help with the section on advanced therapies in Chapter 5.

Introduction

Remarkable progress has been made in the treatment of cancer over recent decades, not only in pharmacological products but also in radiation therapy, surgical procedures and cell and gene therapies. Targeted therapeutics and immunotherapies have radically improved outcomes, and the identification of biomarkers is enabling treatment policies to be tailored to patients who are most likely to respond to particular interventions (precision or personalized medicine). Similarly, systemic and chemotherapy agents are being used for cancer prevention in high-risk individuals.

Clinical trials in oncology involve several challenges for the pharmaceutical industry and academic/public sector organizations that conduct trials. The choice of comparator and trial outcome measures, and the definition of the target patient population, are key considerations. For example, the growing number of treatment options makes the choice of a relevant comparator more difficult, and the background standard of care may change during the course of a trial. Furthermore, regulators, payers (healthcare providers), clinicians and patients often have different expectations that need to be taken into account. The increasing use of molecular profiling (with sensitive and cheaper laboratory tests) has led to the identification of smaller subgroups of patients with defined tumor types, such that large randomized trials may not be feasible and alternative approaches are needed.

Chapter 1 describes the fundamental design features of clinical trials, which provides the framework for Chapters 2–4 focusing on the key attributes of Phase I–III trials of pharmaceutical drugs; Chapter 5 describes trials in surgery, radiation therapy and advanced therapies. The processes and documentation required to set up and conduct a trial are outlined in Chapter 6, and Chapter 7 gives a broad view of how trial data are used, including the importance of publishing, and the role in licensing and market access, as well as the value of real-world evidence. We have focused on clinical trials for treating cancer patients, but the same principles of design, analysis and interpretation apply to preventive, diagnostic and supportive care interventions.

This book provides medical, pharmaceutical and allied health professionals with a concise overview of how contemporary cancer trials are designed and conducted, in order to enhance their ability to critically evaluate published evidence.

1 Fundamental features of clinical trials

This chapter outlines the main features of cancer trials, providing a framework for subsequent chapters. Few new drugs transit the full trajectory from laboratory discovery to clinical practice. Between 2003 and 2011, only 7% of oncology drugs investigated in Phase I–III trials received regulatory approval from the US Food and Drug Administration (FDA).[1] Modern trials present various challenges for the pharmaceutical industry[2,3] and academic and public sector organizations, including administrative burden and high costs. Nevertheless, trials continue to play a central role in research on prevention and treatment.

What is a clinical trial?

A clinical trial is an experimental research study in which some or all of the participants receive an intervention that they would not normally have.

The development of most interventions typically takes 5–15 years from inception through to being recommended for routine care. During this time, several clinical trials provide the main evidence relating to benefit and harms (Figure 1.1). Drugs and some medical devices require a marketing authorization (license) followed by a process of market access that allows them to be provided to the particular patient population (see Chapter 7 and Table 7.1).

Clinical trials are classified into Phases I–IV, with different objectives and designs (Table 1.1). Table 1.2 outlines the main design features of clinical trials, described further in the following sections. These features form the **trial protocol** (the most important document), along with the justification for the study, biological plausibility for the proposed interventions, specific objectives, statements about Phase (I–IV), recruitment processes, safety monitoring and an outline of the main statistical analyses.

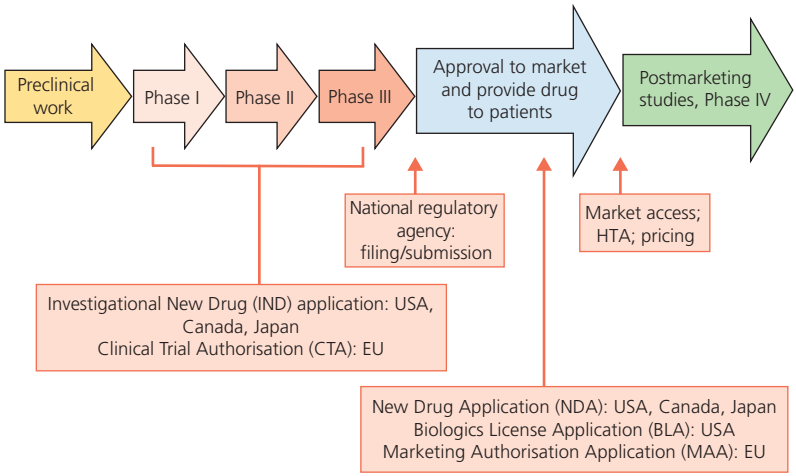


Figure 1.1 The drug development process. Phases may be combined (for example, Phase I/II or II/III). Market authorization (license) and market access are outlined in Chapter 7. HTA, health technology assessment.

Objectives

Each trial will have several objectives (aims or hypotheses). Each objective is typically associated with a clearly defined quantitative outcome measure. The primary objective is meant to inform what happens after the trial if the objective is met, such as a change in practice or further studies. A simple primary objective is to determine whether drug X improves overall survival (OS) in patients with cancer Y compared with standard drug Z. Another example is whether drug A given before surgery (as neoadjuvant therapy) leads to successful complete resection of the tumor.

Secondary objectives (with corresponding outcome measures) provide supporting evidence, although they often also influence decision-making. Examples of secondary objectives include safety and adherence to treatment, and they can be used to provide further knowledge about an intervention (such as whether certain patients benefit more than others) or to find prognostic and predictive markers.

TABLE 1.1

Key features of oncology trials

Phase	Typical number of patients	Design	Primary aims
I	<50	Usually at least one patient cohort and open label Can be first in human	• To show adequate safety • To find a dose (drug or radiation therapy) with tolerable toxicity • To examine biological and pharmacological effects
II	30–100 per group	May be single arm or have several arms, including a comparator (control)	• To obtain a preliminary estimate of efficacy • To further evaluate toxicity • May inform design of Phase III trial
III	Several hundred or thousand patients	Must be randomized and include a comparator group	• To provide definitive evidence on whether a new treatment is better than the control (superiority), similarly effective (equivalence) or not materially worse but with other advantages (non-inferiority)
IV*	Several hundred or thousand patients	Patients from real-world practice Not usually randomized	• To monitor efficacy and safety in the population once the new treatment is used in routine practice • May identify uncommon adverse events not seen in Phase II or III trials

*Also referred to as postmarketing surveillance/pharmacovigilance studies.

TABLE 1.2

Key design features of clinical trials

Feature	Explanation
Eligible patient population	Which patients can be recruited (inclusion and exclusion criteria)
Interventions	Details of the new and comparator treatments, and whether randomization and blinding (placebo) are used
Outcome measures (endpoints)	Primary: one endpoint (sometimes two or three) considered to be the most clinically relevant Secondary: supportive evidence Translational research: biomarkers in blood, tissue or urine samples, or imaging scans
Follow-up	Type, number and timing of clinic visits and assessments, and biological specimen collection required to measure outcomes and monitor patient safety
Sample size	Justification for the number of patients required, usually to demonstrate a statistically significant difference in the primary endpoint

Phase III and many Phase II trials have one of the following general efficacy objectives:

- superiority: the new treatment is more effective than the comparator (or what is expected using current standards of care)
- non-inferiority: the new treatment is not much worse than the comparator
- equivalence: the efficacy of the new treatment is similar to that of the comparator.

Most are superiority or non-inferiority trials. For non-inferiority and equivalence, the new treatment is expected to be safer, cheaper,

easier to administer or have a better health-related quality of life (HRQoL) profile than the comparator (see Figure 4.1).

Patients

Patients are usually enrolled at the point of care, where staff will identify and approach potentially eligible patients. The trial protocol will specify the inclusion criteria (which patients can participate in the trial) and exclusion criteria (which patients should not take part). These criteria aim to ensure that only patients likely to benefit from the new treatment, with minimal harm, are enrolled. Molecular profiling of tumors and identification of particular biomarkers are increasingly used to classify and select patients for trials, particularly for targeted agents.

Major eligibility criteria include:

- a confirmed diagnosis of the cancer(s) of interest by histopathology and/or imaging, along with cancer stage
- adequate fitness to tolerate a new treatment, usually determined by performance status (PS)
- the absence of comorbidities or symptoms that might be exacerbated by the trial interventions (for example, significant abnormal liver or renal function), which could be correlated with PS
- no previous exposure to treatments similar to the experimental treatment (for example, the same drug or class of drug).

PS is a major criterion because it is often correlated with survival and progression; therefore, many trials are restricted to patients with an Eastern Cooperative Oncology Group (ECOG) score of 0–2 or Karnofsky score of 60–100% (Table 1.3).

Clinical trials of cancers in children and teenagers require particular consideration, including drug doses (often different from those used in adults), management of toxicities (close safety monitoring) and measurement of HRQoL (specially designed questionnaires or HRQoL questionnaires for completion by a parent or caregiver rather than the patient). There are specific regulations associated with these trials.

Interventions

New (also referred to as experimental or investigational) interventions are developed for a particular line of treatment: for example, first or second line for solid tumors and many lymphomas, induction or

TABLE 1.3

Commonly used scales to determine a patient's level of fitness to receive trial interventions

ECOG score	Karnofsky scale
0 Fully active	**80–100** Able to carry on normal activity and to work; no special care needed
1 Ambulatory, able to carry out light work, but limited strenuous physical activity	**50–70** Able to live at home with self-care for most personal needs (with varying extent of assistance), but unable to work
2 Ambulatory with self-care but unable to work	**0–40** Unable to care for self; requires equivalent of institutional or hospital care
3 Limited self-care, confined to bed/chair for >50% of waking hours	
4 Completely disabled, cannot self-care, completely confined to bed/chair	

consolidation/intensification therapy for acute leukemias and myeloma, or adjuvant or neoadjuvant therapies. The intervention being evaluated is described in detail in the protocol so that it is correctly and consistently administered across all trial centers, and regulatory agencies and ethics review boards can understand the potential impact on safety.

An experimental pharmaceutical treatment could be a licensed drug that is already available as part of routine care but will be combined with other therapies to form a new regimen or investigated for a different disorder (indication) than specified in its current license, or an unlicensed (novel) agent (referred to as 'first in class' if the mechanism of action is new and unique). Any pharmaceutical drug or micronutrient intended to treat or prevent cancer is called an Investigational Medicinal Product (IMP; for example, in Europe) or an Investigational New Drug (IND; for example, in the USA, Canada and Japan).

For all of the trial drugs to be evaluated, the protocol must include the mode of delivery (for example, oral, intravenous bolus or infusion), dose, frequency and duration. The protocol also sets out any modifications that are allowed to the trial interventions, and the

circumstances in which the intervention should be delayed, reduced or stopped. However, the treating oncologist may override these in the best interests of the patient. Ancillary treatments are also specified in the protocol, including standard supportive care (for example, prophylactic anti-nausea medication).

For radiation therapy (RT), the protocol specifies the dose, number and timing of fractions or brachytherapy treatment. For surgical procedures, the key technical and anatomic aspects are specified but the protocol usually allows surgeons to use their preferred instruments and techniques.

Trials can evaluate a single intervention (monotherapy) or a combination of treatments (for example, several drugs, or drugs with RT or surgery).

Comparator (control)

Randomized Phase III trials and many Phase II studies include a group of patients who receive a comparator treatment, as this is the most reliable way to determine the clinical value of a new intervention. The comparator can be:

- standard of care (SOC): the currently recommended therapy in the clinic, which may differ within and between countries or change while a trial is ongoing
- best supportive care (BSC): any form of palliative management considered appropriate for a patient.

In an open-label trial, patients and the research team will know which treatment is given. However, use of placebo is considered the gold standard in randomized studies, when feasible and ethical, because it minimizes the placebo effect in which patients or clinicians can bias outcome measures in favor of the new therapy and against the control therapy, creating a spurious difference in efficacy.

A placebo is a harmless pill, device or other intervention that has no known therapeutic effect and looks the same as the experimental treatment (and ideally tastes or feels the same). Placebo can be used in addition to SOC or BSC, or on its own.

A double-blind trial is one in which neither the participant nor the research team knows whether the participant is receiving the active treatment or placebo. In a single-blind trial, usually only the participant is unaware of the allocation.

Randomization

Randomization (random allocation) is key to minimizing confounding (baseline differences in patient or tumor characteristics that lead to spurious differences in the endpoints between the experimental and trial interventions) and bias (how patients are selected, allocated to treatment, managed or behave/respond when being assessed might also lead to spurious differences in endpoints). Confounding and bias can make a new treatment appear beneficial when it is not, hide a real benefit or over- or underestimate the magnitude of the benefit.

Randomization ensures that baseline characteristics are comparable between the experimental and control groups, such that any observed differences in outcomes are due to the allocated treatment, and not to confounding or bias. Neither the patient nor the research team has any influence over which treatment is allocated. Computer programs are used to randomize patients; these can be bespoke or commercially or freely available; they include interactive voice recognition systems (telephone randomization) and online services.

Outcome measures (also called endpoints)

These are quantitative measures of the effect of a treatment. They evaluate different aspects of treatment in four categories:

- efficacy
- safety (adverse events [AEs], toxicity)
- adherence to treatment
- patient-reported outcomes (PROs), such as HRQoL.

Efficacy. Phase II and III studies include one primary or two coprimary endpoints used in the primary trial objective. Likewise, the secondary endpoints support the secondary objectives. The scientific value and clinical relevance of different efficacy endpoints are often debated between healthcare professionals and dependent on the cancer type, stage and sometimes type of treatment (such as class of drug). The following endpoints are commonly used.

Tumor response, for solid tumors, is based on whether lesions increase or decrease in size, remain unchanged or new lesions appear during or after treatment – compared with the baseline measurement. This is determined using imaging (radiography, CT, MRI, PET) and established criteria such as the Response Evaluation Criteria in Solid

Tumors (RECIST) in which the change in tumor size is categorized as a complete or partial response (CR or PR), stable disease (SD) or progressive disease, including new lesions.

Tumor response may refer to either the best response seen for each patient or the response at the end of the treatment period (for example, after six cycles). Measures of treatment success are overall response rate (ORR; achieving a CR or PR) or clinical benefit rate (achieving CR, PR or SD). Hematologic malignancies have their own response criteria, based on blood or bone marrow samples, with outcomes such as complete remission/response (no clinical or imaging evidence of disease) and minimal residual disease (a low number of cancer cells remaining in the blood or bone marrow during or after treatment). When evaluating neoadjuvant therapies (particularly for early stage/curable cancers), the primary endpoint could be achieving pathological CR after surgery.

Efficacy biomarkers may be measured in biological samples and used as surrogate endpoints. Examples are cancer antigen 125 (CA 125) for ovarian cancer, Ki67 for breast cancer and prostate-specific antigen (PSA) for prostate cancer. The sensitivity and reliability of new markers need to be ascertained.

Time-to-event outcomes such as OS (Table 1.4) are displayed as Kaplan–Meier curves, which show the cumulative risk of an event over time in each trial arm. Different time-to-event endpoints have strengths and limitations (Table 1.5).

'Hard' versus surrogate endpoints. OS is regarded as a 'hard' endpoint because it has obvious and direct impact. HRQoL is also sometimes considered to be a hard endpoint as it reflects a patient's physical or psychological symptoms. Examples of surrogate endpoints are tumor response, biomarkers and progression-free survival (PFS). Their clinical relevance is less certain because an increase in tumor size by, for example, 30% measured by imaging may not be accompanied by any symptoms nor require immediate treatment, or improvement in the surrogate does not lead to any benefit on a hard endpoint.

A good surrogate endpoint is highly correlated with OS (or other accepted hard outcome), and the treatment effect (for example, hazard ratio [HR]) on the surrogate endpoint should also be highly correlated with the treatment effect on OS. PFS is an accepted surrogate for OS in advanced ovarian and colorectal cancer, and it has been accepted as

TABLE 1.4

Commonly used time-to-event outcome measures

Outcome measure	What defines an event;* all other patients are censored
Overall survival (OS)	• Death from any cause
Disease/relapse-free survival (DFS/RFS)	• First recurrence/relapse of the cancer of interest • First occurrence of secondary malignancies (sometimes excluded from RFS) • Death from any cause
Progression-free survival (PFS) Duration of response (DOR) Duration of clinical benefit (DCB)	• First sign of cancer progression • Death from any cause
Cancer-specific survival	• Death from the cancer of interest
Time to treatment failure (TTF)	• First sign of cancer progression • Death from any cause • Stopped trial treatment
Time to next (anticancer) treatment (TTNT)	• Start of a new anticancer treatment for any reason • Death from any cause

*When a patient experiences several events, only the date of the first one is used. The date of progression or recurrence is based on clinical detection (usually by imaging or a biomarker), not when it has appeared biologically. Any patient who does not have the event of interest is censored, often using the date of their last clinic visit.

Event-free survival (EFS) can have the same definition as DFS or PFS, so it is important to understand what constitutes an event.

DFS and RFS are typically used for early stage (curable) cancers, while PFS, TTF and DOR are often used for advanced disease.

TABLE 1.5

Strengths and limitations of time-to-event endpoints

Strengths	Limitations*
OS	
• Easily defined and precisely measured • Objective • Understood by patients	• Affected by crossover† or subsequent anticancer treatment (treatment effect on OS is diluted or masked) • May require a long follow-up time to observe enough events
PFS, DFS, TTF and TTNT	
• Assessed earlier and has more events than OS, hence requires a smaller and/or quicker trial • Not influenced by crossover† or subsequent anticancer treatment • TTF/TTNT sometimes relevant to patients and healthcare payers	• Affected by assessment bias (especially in open-label studies) • Clinical value uncertain in some cancers (may not be accepted by HTA agencies) • PFS/DFS depend on frequency of assessments • TTNT not easily specified as there may be various reasons to start the next treatment
DOR and DCB	
• Indicates length of time of benefit (understood by patients) • Long duration may reflect early response • May reflect the delay until next line of treatment • Long DOR may reflect a plateau on its Kaplan–Meier curve	• DOR is usually based on only a subset of all patients, i.e. those who have CR or PR • DCB based on only patients who have CR, PR or SD • Baseline characteristics may not be balanced (potential confounding)

*Apply to all the endpoints in the section unless a specific endpoint is indicated.

†Crossover occurs when patients in the control group experience disease progression and they then receive the experimental treatment during the trial.

DCB, duration of clinical benefit; DFS, disease-free survival; DOR, duration of response; HTA, health technology assessment; TTF, time to treatment failure; TTNT, time to next (anticancer) treatment.

the basis for licensing of new treatments in several solid and hematologic malignancies (although mature OS data may still be expected). However, PFS may not be accepted by health technology assessment (HTA) agencies/healthcare payers in some cases.

Safety. AEs may be symptomatic (for example, diarrhea and nausea caused by drugs, or nerve damage and bleeding from surgical procedures) or abnormal biochemical levels in blood or urine (for example, abnormal kidney function tests), which may or may not be accompanied by physical symptoms. The international Common Terminology Criteria for Adverse Events (CTCAE) system is used to classify the severity of AEs (Table 1.6). The following are usually examined in analyses of trials:

- number (and percentage) of patients who experienced an AE of any grade, or focusing on severe events (grades 3–5), sometimes with consideration of how quickly they resolved and whether they were easily treated
- deaths considered to be related to treatment.

Adherence. It is important to quantify how adherent patients are to treatment (sometimes referred to as treatment compliance), because low adherence could reflect unacceptable AEs or compromise efficacy.

TABLE 1.6

Grading of AEs in the CTCAE

Grade	Severity
0	No event
1–2	Relatively minor
3–4	Moderate or serious; may require treatment or hospitalization
5	Death considered to be causally related to treatment

Adherence can be defined in several ways, such as the number of completed cycles of treatment or duration of treatment (in months) per patient, or patients who:

- did not start the new or comparator treatment (or did not receive the allocated procedure in surgical trials)
- started treatment but did not receive the full planned dose (because of dose reductions, interruptions, delays or suspensions)
- stopped the new or comparator treatment earlier than planned, with the reasons recorded.

When all patients receive a SOC background therapy, adherence to this should also be assessed in case patients stop taking it or reduce their dose because of AEs caused by the experimental treatment.

Patient-reported outcomes (Box 1.1) are provided by patients using self-completed questionnaires or face-to-face interviews. They cover HRQoL, which reflects symptoms and wellbeing. The most appropriate PRO questionnaires (instruments) have been developed and validated specifically for patients with cancer or for a particular tumor type,

BOX 1.1

Typical items (domains) that characterize HRQoL and other PROs

- General wellbeing
- Daily living and ability to work
- Physical functioning
- Mental health and psychological functioning (for example, depression, anxiety)
- Social interactions
- Sexual health
- Fatigue
- Pain
- Personal financial costs (for example, cost of travel to clinic, unpaid time off work)
- Satisfaction with treatment

with a focus on key symptoms of that cancer type. The European Organisation for Research and Treatment of Cancer (EORTC) and Functional Assessment of Cancer Therapy (FACT) instruments are commonly used to assess HRQoL. PROs are particularly valued by HTA agencies as they give an indication of the patient experience.

Follow-up

Follow-up refers to the period during which outcome measures are recorded. The duration should allow enough efficacy events to be observed (such as death or recurrence) and other endpoints to be reliably assessed. Longer follow-up is required for trials of early-stage disease in order for recurrences to be identified, whereas follow-up is shorter for late-stage cancer because patients experience progression or die relatively quickly.

How patients are assessed during follow-up depends on the outcome measures. The timing and type of assessments may be the same as in routine practice or involve additional clinic visits and investigations stipulated in the protocol. If, for example, disease progression is expected early on, scans may be required frequently – such as every 2 or 3 months – compared with recurrence for early-stage disease, where 6-monthly or annual scans may be sufficient. The trial protocol will set out the schedule of assessments; they may be more frequent at the start of a trial, but afterward the interval between them lengthens. Telephone assessment may also be used.

Sample size

There are many statistical methods for estimating the number of patients required for Phase II and III trials (Figure 1.2). Estimates of efficacy can be determined using prior evidence, consensus from expert opinion or good clinical judgment. The larger the expected treatment effect, the smaller the trial. Free or commercial software is available to calculate sample sizes.

Translational research

Biological specimens (tumor tissue, bone marrow, blood or urine) for translational research linked to the trial may be obtained from excess samples taken as part of routine practice or specifically for the study.

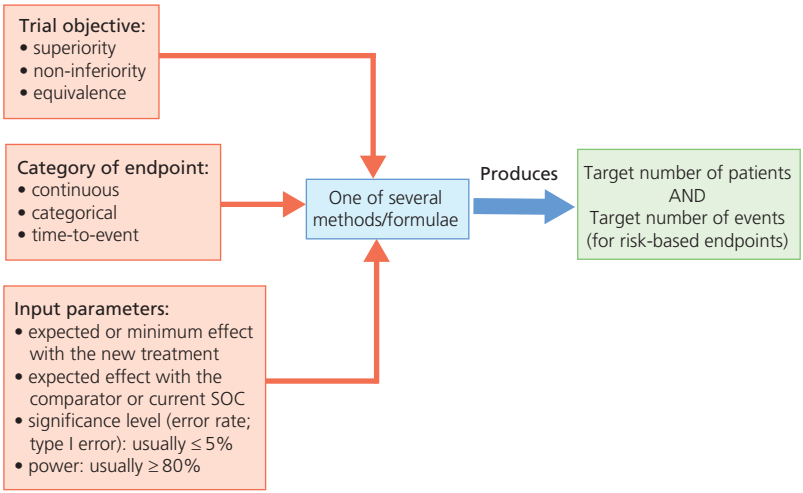


Figure 1.2 Information needed for a typical sample size estimation for Phase II and III trials. Significance level: the chance of (falsely) concluding that a new treatment is more effective than a comparator when it is truly not. Power: the chance of finding a treatment benefit when it truly exists, and that the effect is statistically significant at the prespecified significance level.

Translational research questions may be formulated when the trial is designed (as secondary objectives that potentially influence the main trial findings or exploratory objectives) or after it has finished. These analyses can provide further insight into the mechanism of action of the new treatment or identify prognostic or predictive biomarkers (see page 68). High-quality scans may also be used in translational research.

Next-generation sequencing (NGS), based on whole-genome or exome sequencing, and other biomarker technologies increase knowledge of cancer biology, allowing highly effective therapies to be developed. Several trials include a biomarker as a key eligibility criterion, which has led to approved targeted drugs for patients with certain tumor genetic mutations, for example *EGFR*, *BRAF* and *RAS* mutations, *ALK* and *ROS1* rearrangements and *NTRK* gene fusions; also PD-L1 expression for immunotherapies. High-quality multi-gene NGS panels are required, which may already be part of local clinical practice. Although NGS has, historically, nearly always been performed on tumor tissue, liquid biopsies (blood) have now been approved.

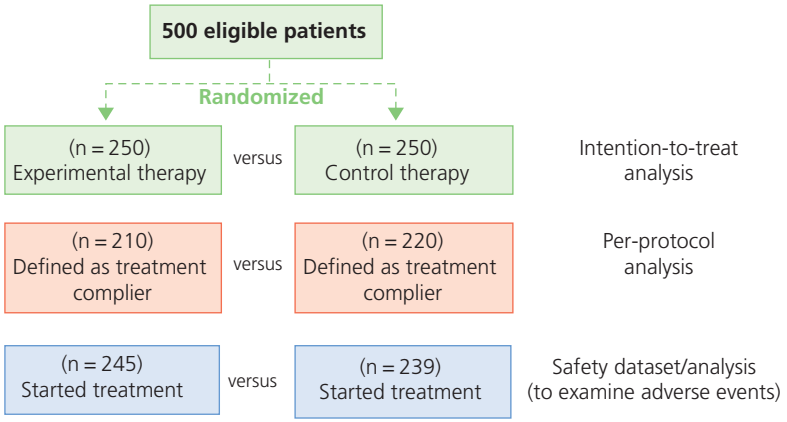


Figure 1.3 The main types of analyses of a clinical trial. A 'treatment complier' indicates adherence, and is predefined in the trial protocol (for example, started trial treatment, finished trial treatments as intended, or had ≥80% of the planned dose). For ITT analyses, patients are analyzed according to the group they were allocated to, regardless of what (and how much) treatment they actually received.

Detecting mutations in circulating plasma or blood offers an easier way of assisting with the diagnosis and monitoring of patients.

The conventional premise of cancer therapy is to treat a specific cancer type, usually based on the anatomic location of the primary tumor. However, the field referred to as 'tumor agnostics' involves treating a specific biomarker (mutation) with one targeted drug, across a range of cancer types combined (see basket trials in Chapter 3). Many mutations are uncommon such that only single-arm trials are feasible, with tumor response and duration of response as key endpoints. Drugs have been given early approvals by the FDA and European Medicines Agency (EMA) in the absence of randomized trials (examples are larotrectinib and entrectinib for solid tumors with *NTRK* gene fusions). Payers and other decision-makers may need to accept that these particular types of drugs might often only be evaluated using high-quality single-arm trials together with real-world data.

Statistical analyses

Figure 1.3 outlines the main types of analyses performed for clinical trials. Intention-to-treat (ITT) analysis is the gold standard and ensures that the comparison of outcomes measures is based on balanced baseline characteristics (achieved by randomization), but it can sometimes underestimate the treatment effect when many patients have insufficient adherence to the treatments. Per protocol analyses are especially useful for non-inferiority trials and should be predefined and justified in the protocol.

Key points – fundamental features of clinical trials

- A clinical trial is an experimental study to evaluate an intervention in patients.
- Trials are generally classified into four Phases (I–IV), depending on the objectives.
- The trial protocol is a key document that specifies how the trial was designed and how it will be conducted and analyzed.
- The main design features of clinical trials are patient eligibility (definition of the patient population and number required), specification of the trial interventions including control (comparator) therapies, the outcome measures and the follow-up schedule.

References

1. Hay M, Thomas DW, Craighead JL et al. Clinical development success rates for investigational drugs. *Nat Biotechnol* 2014;32:40–51.

2. Moscicki RA, Tandon PK. Drug-development challenges for small biopharmaceutical companies. *N Engl J Med* 2017;376:469–74.

3. Rosenblatt M. The large pharmaceutical company perspective. *N Engl J Med* 2017;376:52–60.

2 Phase I trials

Extensive preclinical laboratory studies (in vitro and animal studies) are performed before an experimental drug (or new combination of drugs) is given to humans, to develop an understanding of toxicity and the pharmacological profile.[1] A variety of biological materials are used in preclinical studies to understand how a drug works, including cell lines derived from human solid tumors, and liver and blood cells. A drug that has little effect on preclinical efficacy markers is unlikely to progress into clinical trials in humans.

Purpose

Phase I trials are used to evaluate the pharmacological actions and safety of an unlicensed anticancer drug, a new treatment combination (using licensed or unlicensed drugs) or a licensed treatment in a new indication (for example, different tumor type or line of treatment).[2] New ways of delivering RT, such as intensity-modulated RT (IMRT), are also explored in Phase I trials. When a drug has not been tested on humans previously, Phase I trials are called first in human studies.

Phase I studies are a precursor to larger-scale Phase II and III studies. They are often dose-escalation (dose-ranging or dose-finding) trials, in which several different doses of a drug or RT are administered and AEs (toxicities) monitored. One of the aims is to determine the maximum tolerated dose (MTD) that has an acceptable AE profile; this dose is likely to be used in subsequent studies. One or more Phase I studies may be conducted with an experimental drug to evaluate how it is best administered: for example, whether it should be taken with or without food, the optimal time of day (morning or night), the dose and optimal frequency (for example, every day, every 3 days or, in the case of some monoclonal antibodies, every 1–3 weeks).

Phase I studies may be preceded by a Phase 0 trial in a few patients (usually <10), to determine whether the drug reaches cancer cells and how these cells change, the route of elimination and the pharmacological

profile. The dose of the drug is typically so low that it is unlikely to have any effect on the tumor or cause toxicities. Participants are monitored closely using multiple imaging scans and blood samples. Phase 0 trials are a type of first in human study.

Participants

Phase I trials of most new treatments for other disorders are conducted in healthy volunteers, whereas in oncology these studies usually involve patients with cancer, because it would be inappropriate to expose healthy people to drugs that have expected and potentially serious toxicities. Patients with multiple tumor types may be recruited, or the trial may be restricted to a single tumor type.

Participants in Phase I studies have usually received standard treatment and their tumor has become refractory (no longer responds to treatment) or progressed, and they therefore have a poor prognosis. Patients with advanced disease may tolerate treatment less well than patients at earlier stages of disease, meaning that the most tolerable dose observed in Phase I trials could be lower than the most effective dose.

Trial designs

Phase I trials may simply involve giving the experimental treatment (at the same dose) to a relatively small number of patients and examining safety and efficacy. There are also a wide variety of dose-escalation designs.[2–4]

A group of patients (a dose cohort) receives the starting (lowest) dose of the experimental therapy. If toxicity is acceptable, the next cohort receives the next highest dose; this process continues until the MTD is found or the maximum planned dose is reached. Many Phase I trials are standalone, but they can also be included as the first stage in Phase II study protocols (Phase I/II trials). Phase I oncology trials are mostly open label (no blinding or placebo).

Toxicities of most concern are those considered to be causally related to the experimental treatment: dose-limiting toxicity (DLT; Box 2.1). Many DLTs are symptomatic.

BOX 2.1

Dose-limiting toxicities

- A DLT is an adverse reaction considered to be unacceptable – usually a grade 3–5 (CTCAE) event, and it may require treatment, hospitalization or withdrawal of the experimental treatment
- DLTs typically occur relatively soon after starting the experimental treatment (for example, during the first 28 days or first one or two cycles of chemotherapy) and are thought to be causally related to the treatment
- The investigators pre-define events that constitute a DLT and the relevant time period in the protocol
- A toxicity that is causally related to the drug should be distinguished from symptoms of the disease or progression
- The MTD is the highest dose that has an acceptable proportion (number) of DLTs, and is therefore used in further studies

Dose-escalation trials have either rule- or model-based designs (Table 2.1).

Rule-based designs, particularly the 3+3 design (Figure 2.1), are most widely used because they are easy to understand and implement – they are acceptable if only two or three dose levels are being evaluated.[5] However, these designs are generally inefficient if more than three doses are planned; many patients may be required to reach the MTD because of a lack of DLTs in the lower-dose cohorts, which provides little information about the dose–toxicity relationship. Also, the assessment of toxicity using DLTs is only made for each drug dose, essentially ignoring the DLTs seen at other doses.

Model-based designs. The MTD is typically reached more quickly using model-based designs when there are several planned doses. Fewer patients are required than in the 3+3 design, and all dose cohorts are used to describe the dose–toxicity relationship. DLTs can be assessed after each patient is treated and evaluated, so that the dose–toxicity model can be updated in real time to determine the dose for the next patient(s). The continual reassessment method (CRM) is commonly used (Figure 2.2).[6]

TABLE 2.1

Dose-escalation trial designs

Rule based	Model based
• Assumes nothing about the shape of the relationship between dose and toxicity	• A mathematical relationship between dose and toxicity is assumed (a flattened S-shaped curve)
• A + B design: 'A' patients are recruited into a dose cohort, which may then be expanded by another 'B' patients (A and B are numbers); 3 + 3 design is the most commonly used (Figure 2.1)	• The initial shape of the relationship is estimated and then modified after each patient or patient group has been treated and assessed, using statistical modeling (Bayesian methods or regression analyses)
• The DLT rate that is considered unacceptable is used in the design (e.g. more than one-third of patients for 3 + 3 designs)	• Low doses that are not expected to have DLTs can be missed, and some dose cohorts comprise only 1 patient
• Accelerated titration designs use 1 patient per dose from the starting dose, but if DLTs or moderate grade toxicities are seen, a 3 + 3 design begins	• CRM is most commonly used (Figure 2.2)

The starting and subsequent doses of a drug (or RT) is based on the expected toxic and effective doses from preclinical studies, and local regulatory requirements. For example, the FDA requires evidence from mammalian species.[7] The initial dose is usually one-tenth of the dose that is associated with severe toxicity in 10% of rodents, or one-sixth of the highest non-severe toxic dose in mammals other than rodents, using an appropriate species.

If the investigational treatment is already licensed (for another disorder or cancer type), the starting dose could be the same as that currently used, or lower if there is concern over toxicity, especially if the drug or RT is to be combined with other therapies. Several approaches

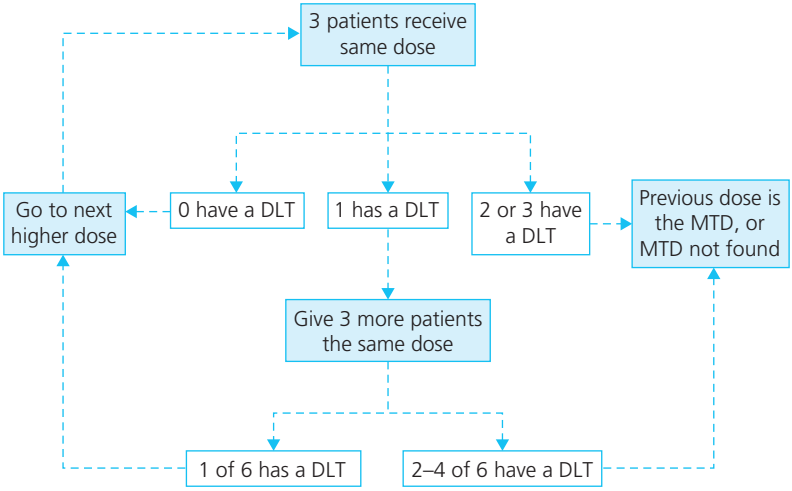


Figure 2.1 A '3+3' Phase I trial design: the dose is increased until the maximum planned dose or MTD is reached.

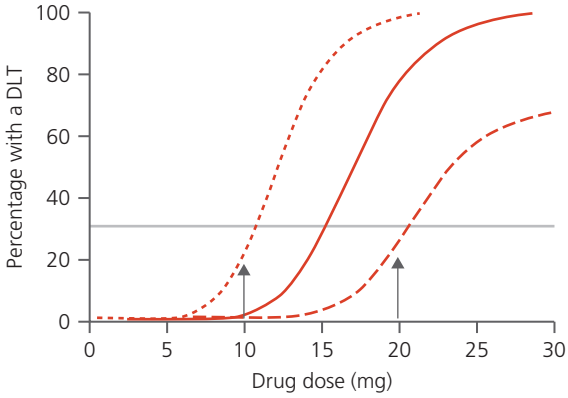


Figure 2.2 Continual reassessment method. A DLT rate greater than 30% is considered unacceptable (horizontal line). The solid (middle) curve is the estimated initial dose-DLT relationship. If the starting dose is 15 mg and there is no DLT, a remodeled curve (right dashed curve) indicates that the next dose could be 20 mg. If a DLT occurs at 15 mg, the remodeled curve (left dotted curve) indicates the next dose could be 10 mg. This is because the expected DLT rate is less than 30% in both cases (arrows).

are used to determine subsequent doses. They may come from preclinical studies or other early-phase trials. Doses may be fixed by the physical size of the tablet or capsule or constrained by the type of RT. Alternatively, a mathematical Fibonacci sequence, or modified version of it, may be used in which each dose is the sum of the two previous doses.

Follow-up for each dose cohort is typically the first treatment cycle (for example, 3 or 4 weeks) or 28 days after starting the experimental treatment, to allow DLTs and other toxicities to be observed before the next higher dose is administered to the next cohort. A longer assessment period (for example, the first two cycles) may be required sometimes, particularly for novel agents which have limited prior evidence, and new therapeutic combinations.

In Phase I trials of RT, some AEs only occur after several months but it is rarely feasible to allow a long assessment time before treating the next patient cohort. A model-based design (time-to-event CRM) is useful in this situation. Late events are not classified as DLTs but are reported as part of the whole toxicity profile.

Further considerations. Occasionally, few or no DLTs are seen across all dose cohorts so the MTD is not found but the highest dose shows sufficient evidence of biological activity (using markers of response). Once the MTD has been reached, it is good practice to recruit a further group of 6–12 patients to receive the same dose, to obtain more data on toxicity and efficacy markers.

Many novel pharmaceutical agents are expected to be combined with standard treatments (for example, chemotherapy and RT), which may require the recommended dose of the standard treatment to be reduced to minimize toxicity. In addition, drug–drug interactions may lead to more severe AEs or more DLTs than expected with either drug on its own, although there may also be positive synergistic effects on markers of response (efficacy). Model-based designs are well suited for evaluating changing doses of two drugs.

Dose de-escalation studies. Occasionally, dose de-escalation trials are conducted if there is prior evidence of toxicity when using the therapy in a different disorder. The first cohort of patients receives the highest

planned dose; further cohorts receive lower doses only if DLTs are seen. This approach would have to be justified in regulatory and ethics submissions but can result in a quicker trial than a dose-escalation study.

Outcome measures

Outcomes and biological parameters are usually measured at multiple time points for each patient; the frequency depends on how quickly these effects are expected to occur and change. Pharmacodynamics (PD) and pharmacokinetics (PK) are key outcomes in a Phase I trial. Markers of response (efficacy) may also be included.

Adverse events are an essential outcome, especially when the primary objective is to determine the MTD. The CTCAE system is used to classify and grade AEs (see Table 1.6). Many types of AE occur and are often minor and transient, so investigators tend to focus on grade 3 or higher events.

Pharmacodynamics show how the drug affects the body using physical or biological measures, including markers of activity assessed using imaging scans, or molecular changes (for example, tumor target inhibition or activation). PD measures can also include blood pressure, body temperature, heart rate, respiratory rate, liver and renal function, cardiac tests and anything else considered necessary given the cancer type and type of experimental therapy. Toxicities can be considered a type of PD.

Pharmacokinetics show how the body deals with the drug, with assessment of: drug uptake (absorption), distribution, metabolism and excretion. PK measures come from summaries of plasma concentration–time curves for each patient (Figure 2.3).

Efficacy outcomes. Although Phase I trials are not designed to evaluate efficacy reliably, outcome measures such as tumor response, PFS and established blood or tissue markers (for example, CA 125, Ki67 and PSA) may be collected. Striking effects have occasionally been seen and used to justify an expedited clinical trial program.

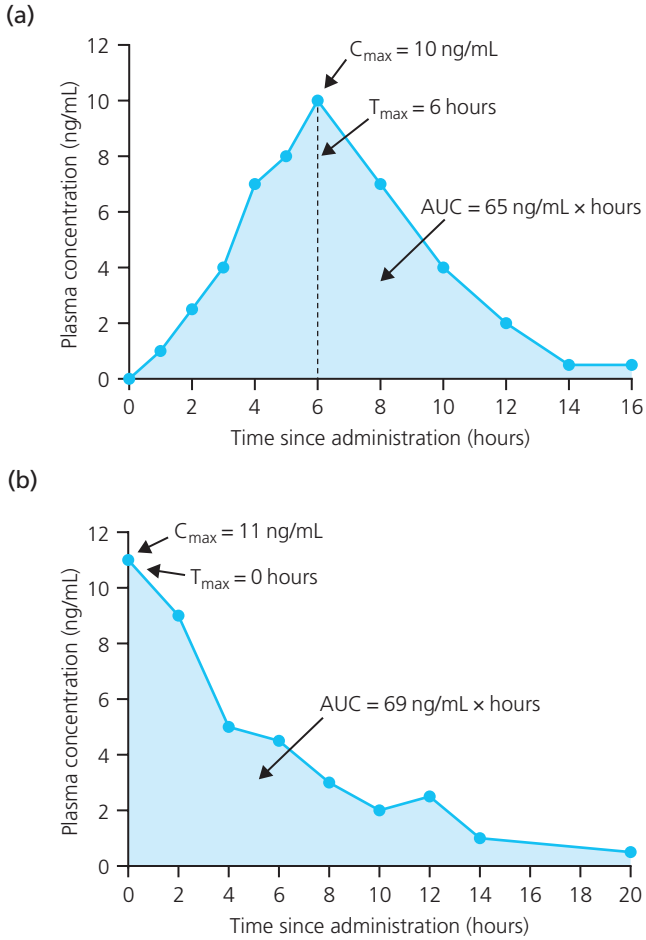


Figure 2.3 PK measures from plasma concentration–time curves following (a) oral and (b) intravenous drug administration. AUC, area under the curve; C_{max}, maximum plasma concentration; T_{max}, time at which C_{max} occurs. Other PK measures are: terminal half-life ($t_{1/2}$), the time taken for the plasma concentration to decrease by 50% later in the curve (drug elimination); clearance (CL), the rate at which the drug is removed from the plasma as it is metabolized/excreted (CL = dose/AUC); volume of distribution (V), the amount of drug in the body divided by the plasma concentration; and bioavailability (F), the percentage of the administered dose that gets into the systemic circulation (intravenous drugs have F = 100%).

Minimum effective dose. Vaccines and some targeted agents are intended to have relatively few AEs, so the concepts of DLT and MTD may be less relevant. The main objective of Phase I trials for these treatments could be to find the *lowest* dose that has a clinically important effect on biological markers of activity (tumor response) – the minimum effective dose (MED) or optimum biological dose. Therefore, efficacy markers are the primary endpoints, rather than AEs. Toxicity is still monitored closely but the MED is equally or more important when determining which dose is carried forward to the next stage of development.

Sample size

The number of patients required can only be approximated; the actual number enrolled will depend on how many doses are evaluated, the DLTs that occur and the design (see Table 2.1). The number of planned doses is often known, so the maximum number of patients that could potentially be recruited can be determined based on conservative assumptions. The majority of Phase I trials are small (typically <50 patients), although trials based on several tumor types may be larger (>100 patients).

Conduct

Health professionals who administer the experimental treatments in Phase I trials must be experienced in dealing with unexpected or serious toxicities. Therefore, some Phase I trials are conducted in designated (expert) clinical trials facilities, and patients are admitted so that they can be closely monitored. However, if there is already evidence of the drug's safety profile, trial participants can be seen as outpatients. Patients have regular investigations, such as blood tests, imaging (for example, CT, MRI, PET, ultrasonography) and physical examinations. Hospital-based trial units may have en suite laboratory, cardiac and metabolic monitoring facilities, access to high-quality imaging, and pressure-controlled rooms (to minimize exposure to contaminated air).

Interpreting and reporting results

Table 2.2 shows an example of a dose-finding trial.[8] The trial results and report tend to be descriptive and should include:

- a summary of the patient and main tumor characteristics
- a summary of all AEs and their severity and likely relationship to the study drug (causality), particularly grade 3–5 events
- a table outlining how the MTD was determined, listing all DLTs
- a description of grade 5 events (treatment-related deaths) including consideration of whether the underlying cancer was a contributing cause
- a summary of adherence to the experimental treatment, including how many patients stopped early and why; and adherence to other anticancer treatments given at the same time
- summary statistics (for example, mean and standard deviation) for PK and PD measures
- any evidence of efficacy/activity, for example, waterfall plots (Figure 2.4).

The aggregated information above, particularly on safety, helps to determine whether to investigate the intervention further, and at which dose.

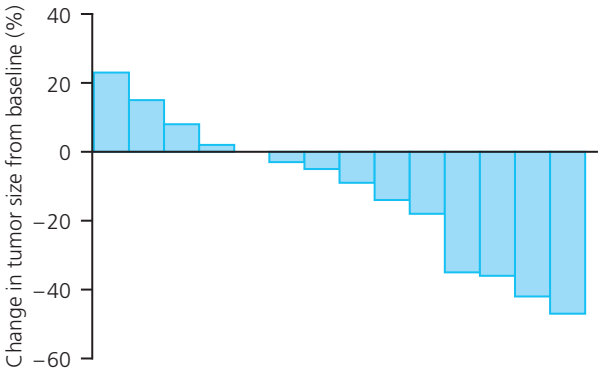


Figure 2.4 A waterfall plot for 14 patients. The tumor size at the end of treatment (or other time point) is compared with the baseline value (before treatment). The percentage change in tumor size for each patient is displayed in rank order. Negative values indicate that the tumor has shrunk. Highly effective treatments will have few positive values and many large negative ones.

TABLE 2.2

An example of a Phase I dose-finding trial, investigating capecitabine in combination with oxaliplatin

Patients: Advanced solid tumors unresponsive to or untreatable with standard chemotherapy

Investigational drug: Oral capecitabine (twice daily, days 1–14), every 21-day cycle

Concurrent therapies: Oxaliplatin (fixed dose 130 mg/m^2 on day 1)

DLT assessment period: First treatment cycle (21 days)

Design: 3 + 3 dose escalation

Results:

Cohort	Capecitabine dose (mg/m^2 twice daily)	No. of patients	Patients with a DLT
1	500	3	None
2	825	3	None
3	1000	3	Grade 3 diarrhea
4	1000	3	None
5	1250	3	Grade 3 diarrhea with thrombocytopenia
6	1250	3	Grade 4 diarrhea with neutropenia
7	1000	3	None

- Seeing 1 DLT in cohort 3 meant that another 3 patients were given 1000 mg/m^2
- No other DLTs were observed, so the next highest dose was given
- At 1250 mg/m^2, 2 of 6 patients had a DLT, which was considered unacceptable
- The final cohort was given 1000 mg/m^2; which became the MTD
- If this trial were conducted today, it might benefit from using a model-based design: fewer dose cohorts and/or fewer patients

From Díaz-Rubio et al. 2002.[8]

Key points – Phase I trials

- Phase I trials provide an early assessment of the safety of a new anticancer treatment or combination, or an existing treatment in a new indication (tumor type or line of treatment).
- Most Phase I trials aim to identify DLTs to determine the MTD. Most are dose-escalation studies in which successive cohorts of patients receive a higher dose of treatment, depending on the toxicities seen at the preceding dose.
- Simple (rule-based) designs can be used to determine the MTD. Model-based designs should be more efficient when several doses are planned.
- Reports of Phase I trials should provide clear information on the pharmacological properties of a new therapy, and the frequency, severity and suspected causality of AEs.

References

1. Muller PY, Milton MN. The determination and interpretation of the therapeutic index in drug development. *Nat Rev Drug Discov* 2012;11:751–61.

2. Eisenhauer EA, Twelves C, Buyse M. *Phase I Cancer Clinical Trials: A Practical Guide*. Oxford: Oxford University Press, 2006.

3. Le Tourneau C, Lee JJ, Siu LL. Dose escalation methods in Phase I cancer clinical trials. *J Natl Cancer Inst* 2009;101:708–20.

4. Cook N, Hansen AR, Siu LL, Abdul Razak AR. Early phase clinical trials to identify optimal dosing and safety. *Mol Oncol* 2015;9:997–1007.

5. Wheeler GM, Sweeting MJ, Mander AP. AplusB: a web application for investigating A + B designs for Phase I cancer clinical trials. *PLoS One* 2016;11:e0159026.

6. Wheeler GM, Mander AP, Bedding A et al. How to design a dose-finding study using the continual reassessment method. *BMC Med Res Methodol* 2019;19:18.

7. US Food and Drug Administration. *Good Review Practice: Clinical Review of Investigational New Drug Applications*, 2013. www.fda.gov/media/87621/download, last accessed 2 June 2020.

8. Díaz-Rubio E, Evans TR, Tabemero J et al. Capecitabine (Xeloda) in combination with oxaliplatin: a phase I, dose-escalation study in patients with advanced or metastatic solid tumors. *Ann Oncol* 2002;13:558–65.

3 Phase II trials

Phase II trials provide preliminary evidence of efficacy and further data on the safety profile of a new intervention, often following Phase I studies, and are primarily conducted to determine whether a treatment is worth investigating in a large Phase III trial. The majority of treatments examined in Phase II studies are pharmaceutical agents, with fewer trials on RT and fewer still on surgery. Having a good Phase II design should increase the likelihood of a Phase III trial achieving its objectives.

While the results of a Phase II study often inform the design of a Phase III trial, they can change clinical practice directly – most often in rare cancers or subtypes for which Phase III trials are not feasible. In addition, Phase II studies that generate impressive efficacy data for cancers with high unmet need may be sufficient to support licensing by a regulatory authority as part of an early access or fast track scheme. This approval may be contingent on the subsequent submission of survival data. However, obtaining market access approval based on Phase II data alone can be challenging.

Participants

Whereas Phase I trials often recruit patients with advanced cancer who have received one or more lines of therapies, Phase II studies aim to enroll patients from the target population of interest. Although many studies involve patients with advanced cancer, they can also explore treatments for early-stage cancer, sometimes with a focus on AEs, efficacy biomarkers and HRQoL.

Common designs

A wide variety of Phase II designs are available and this flexibility can increase the likelihood of identifying an effective new intervention (Figure 3.1).[1–3] There is no standard definition of a Phase II trial, or

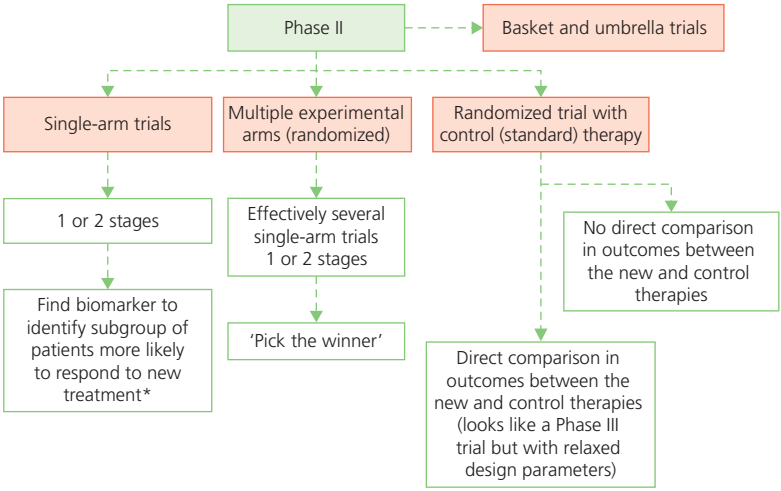


Figure 3.1 Phase II trial designs (*can also be applied to randomized designs).

what distinguishes it from Phase III, but the following are common examples:

- feasibility study to assess recruitment or ability to deliver a new intervention (including patient acceptance)
- single-arm study without a comparator (control) treatment[4]
- use of a surrogate endpoint, including biomarkers, as the primary outcome measure[4,5]
- the sample size method uses one-sided statistical significance tests with $p>0.05$.[4]

Examples of the last three of these points are shown in Figure 3.2.

Phase II studies may be designated 'Phase IIa' or 'Phase IIb'. Phase IIa usually indicates a single-arm trial designed to show initial evidence of clinical efficacy ('proof of concept'). Phase IIb indicates a more reliable assessment of efficacy in a larger number of patients, often randomized with a control therapy group.

Single-arm trials, in which all patients receive the experimental treatment, should be the simplest, quickest and cheapest to conduct. The efficacy outcomes are compared with those from past patients

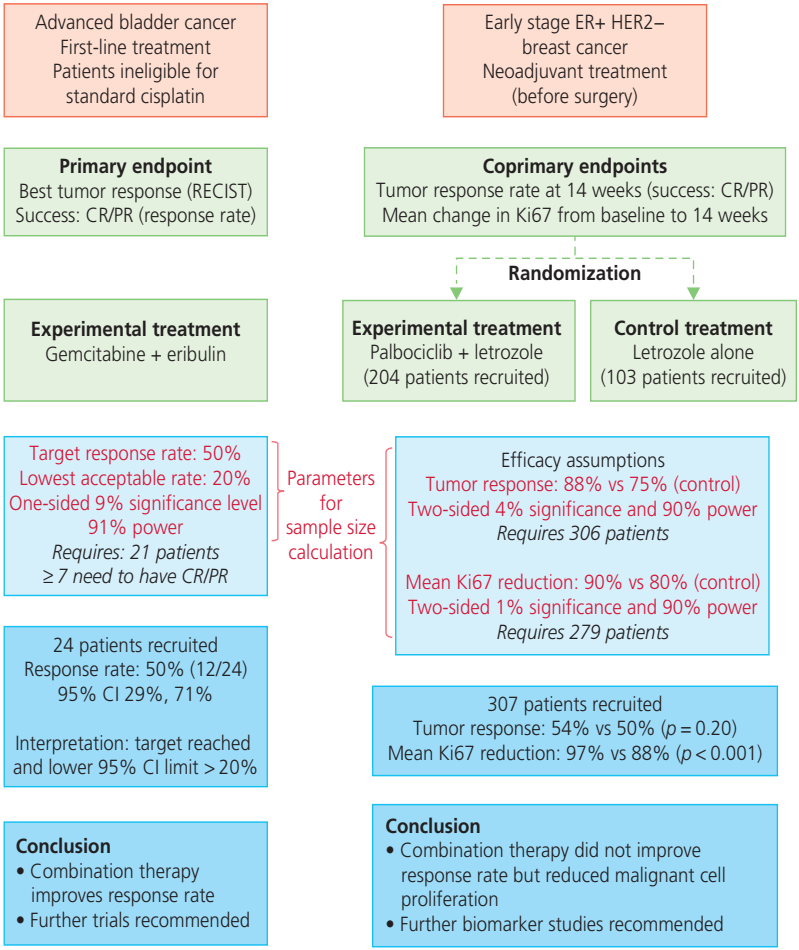


Figure 3.2 Examples of a single-arm trial (left) and a randomized controlled trial (right).[4,5] CI, confidence interval; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; vs, versus.

who were treated with SOC (historic controls). This design has several uses (Box 3.1).[6,7] However, a major criticism of single-arm studies is that, with advances in the management and background treatment of patients, a contemporary control group has better outcomes than historic controls. The effect of an experimental therapy may therefore be overestimated when compared with historic controls, and Phase III

BOX 3.1

Situations in which single-arm Phase II trials may be valuable

- As a proof of concept study in a relatively small number of patients (first assessment of activity/efficacy), for example, for a new class of drug or new combination of treatments
- Rare tumor type or subtype, or small patient group with a specific molecular profile where a randomized trial is difficult
- Standard anticancer treatments are associated with poor outcomes but the efficacy endpoints seen with the experimental treatment show extraordinarily large benefit
- The new treatment has a novel mechanism of action
- The trial incorporates a biomarker that validates a mechanism of action, allows enrichment of patients to produce a large treatment effect or is a novel biomarker that provides new insight into how patients respond to the experimental treatment

Adapted from Cannistra, 2009.[6]

trials are more likely to fail to meet their objectives if the conclusions from the single-arm trial are too optimistic. However, if the management and treatment of historic controls is known to have been relatively stable over time, they could act as a valid comparator for a single-arm trial.

A two-stage single-arm trial may be conducted if a new drug is expected to have serious toxicities or is expensive. The experimental therapy is given to a few patients first; if the efficacy outcomes indicate inactivity or lack of sufficient response, recruitment stops early. Otherwise the trial continues.

Randomized (multiple-arm) trials are generally preferred because randomization minimizes confounding and bias (see page 16).[3,8]

Pick the winner. With this design, patients are randomly allocated to two or more experimental treatment arms, which could be for different drugs, combinations of drugs or doses of the same drug. The primary aim is to see which group has the highest response or most

favorable markers of activity/efficacy, rather than to directly compare the experimental arms. The chosen arm (sometimes two arms) is likely to become the treatment arm for future trials, taking toxicity into account. Two-stage designs can also be used.

A randomized controlled trial involves randomly allocating patients to receive either the experimental or control therapy (for example, current SOC, with or without placebo), using a 1:1 or 2:1 allocation ratio. There are two general approaches. One in which the primary analysis is based on the experimental group alone (effectively a single-arm study) and the control group is used largely to check the efficacy assumption associated with the SOC treatment used in the sample size calculation (see Figure 1.2). With the second approach, the primary analysis involves a direct statistical comparison of the results between the experimental and control groups (same principle as for randomized Phase III trials); this uses the control arm data more efficiently than the first approach but requires a larger study.

Use of a placebo can strengthen Phase II trial results and conclusions, especially when there are subjective endpoints (for example, tumor response and disease progression) that can be influenced by the lack of blinding in open-label trials. However, the cost of producing and distributing placebo in a Phase II trial may outweigh the scientific benefits, and investigators accept the potential for bias, with the intention of using placebo in a subsequent Phase III study.

Outcome measures

Many Phase II trials have a primary objective and corresponding endpoint, but other secondary efficacy endpoints may have similar importance. Therefore, while Phase II studies provide an initial evaluation to identify potential benefit, they can also be used to identify the outcomes most likely to be improved, which then become the primary focus in subsequent Phase III trials. Ideal efficacy endpoints can be measured relatively quickly and are clinically meaningful (see Tables 1.4 and 1.5). Examples are shown in Figure 3.2. Much data may be collected (for example, vital signs and blood and urine biochemical values) to identify safety issues.

OS is the gold standard endpoint in oncology, but while it can be measured in patients with advanced disease and a poor prognosis (survival less than 12 months), the required follow-up is too long for

most Phase II studies and certainly for early-stage cancers. Although PFS is increasingly preferred over tumor response as a primary endpoint in Phase II trials of solid tumors, tumor response can still provide useful information. This is particularly the case for advanced disease when rapid disease progression is expected, so that even maintaining stable disease is considered a positive outcome. In several solid tumor types, PFS seems to be better (but not necessarily *strongly*) correlated with OS than tumor response.

PRO data are not always collected in Phase II trials because of the resources required. Also, HRQoL data can have limited scientific value in (small) single-arm Phase II studies. However, in some circumstances, such as in rare tumor types, PRO data may be collected to get an idea of how they change over time. Randomized Phase II studies increasingly collect PRO data to identify which symptoms are influenced by a new treatment, and this information can be used to design a Phase III trial.

Endpoints that could be biased in open-label trials (for example, tumor response or progression) could benefit from having a central independent review of images, where the reviewer is unaware of the treatment allocation and clinical outcomes of the patients.

Sample size

Occasionally, there is no formal sample size justification in the trial protocol, and investigators aim to recruit what they consider to be a sufficient (usually small) number of patients. However, most Phase II studies employ one of several sample size methods, many of which use (1) the expected (target) measure of efficacy using the experimental intervention, or the minimum effect that would be considered clinically important, and (2) the effect in patients given SOC, or an effect considered to be the lowest acceptable response (see Figure 1.2). These are estimated by the investigators using prior evidence or experience. The larger the difference in these two measures, the smaller the trial required. Figure 3.2 shows two examples.

Other parameters used to calculate sample size (see Figure 1.2) are usually less stringent at Phase II than at Phase III (for example, one-sided significance level of 5–20%). For example, in the single-arm trial in Figure 3.2, the significance level is 9%. Two-sided levels are used for Phase III trials to allow for the new treatment to be better or

worse than the control and are therefore the most conservative, while one-sided levels only allow for the new therapy to be more effective. Using significance values above 5%, one-sided levels or both mean that fewer patients are required for the trial compared with a typical Phase III study (two-sided significance level ≤5%).

Trials using biomarkers

Biomarkers can support the development of targeted therapies designed to maximize benefit in selected patients.[8] Biomarkers measured at the end of the trial (retrospectively) can be correlated with outcomes such as tumor response or PFS. Alternatively, biomarkers can be used to direct the trial treatments when measured prospectively as part of the design (Figure 3.3).[9,10] NGS can identify a range of tumor gene expressions/mutations in tumor and blood samples, while body scans can be used to find imaging markers.

Basket trials involve testing a single drug targeted at a specific marker (for example, a mutation) in several tumor types or subtypes (one drug in several cancers).

Umbrella trials categorize patients with a single tumor type into different groups according to a biomarker or mutation; each group receives a targeted agent for that biomarker (several drugs in one cancer).

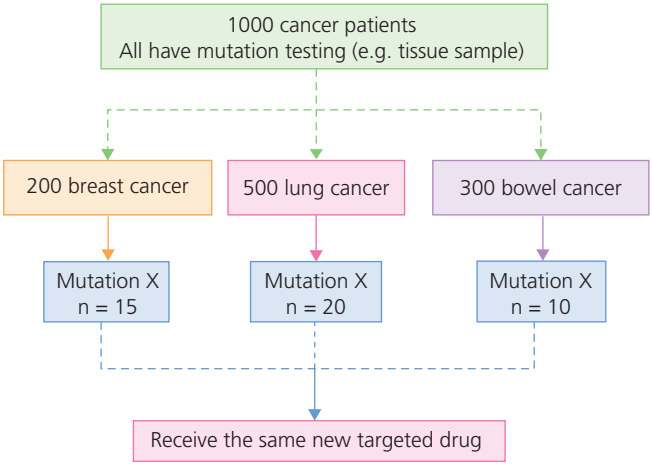
Trial analyses and interpretation

The baseline characteristics of the trial patients are summarized in a table, including age and sex distribution, disease stage and other clinical, histopathological and imaging features relevant to the tumor type.

Single-arm and 'pick the winner' randomized trials require mainly descriptive analyses, in which the efficacy endpoint data are summarized for each group separately. Tumor response data are usually presented as the proportions of patients whose best response was CR or PR. A waterfall plot can be used to demonstrate changes in tumor size from baseline to the end of treatment for each patient (see Figure 2.4). Figure 3.2 illustrates how the design parameters influence the interpretation of results.

HRQoL can be presented as scatter plots, box plots or bar charts that show mean values over time to see whether specific symptoms remain stable, improve or deteriorate (see the example in Figure 4.6).

(a) Basket trial: one drug – several tumor types (or subtypes)



(b) Umbrella trial: one cancer – several drugs

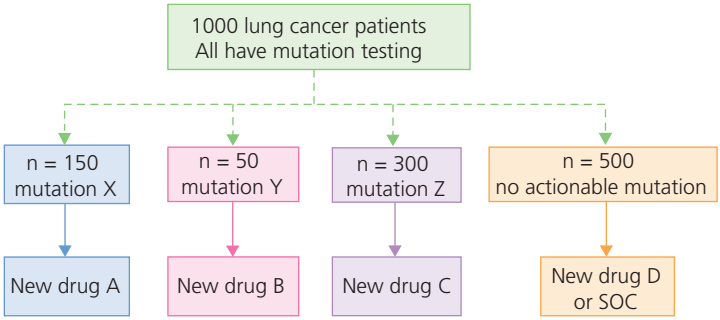


Figure 3.3 Typical designs of (a) basket and (b) umbrella trials which incorporate biomarkers in the design (often using NGS gene panels). These designs have several arms (with or without randomization) and they may or may not have a control group. A basket design was used in the trial of vemurafenib for treating *BRAF* V600-positive tumors;[9] the umbrella design was used in the BATTLE-2 trial in advanced non-small-cell lung cancer.[10] Adaptative randomization can also be used in umbrella trials. For example, the first 100 patients are randomized equally to drugs A–D and efficacy assessed; the remaining patients are then allocated to the drug with the highest chance of response given the mutation result.

Time-to-event endpoints such as PFS and OS are presented as Kaplan–Meier curves, event rates at prespecified time points, and median values (see examples in Figure 4.4 and Box 4.2).

In the single-arm trial in Figure 3.2, the tumor response rate using the combination therapy reached the target, so this study met its objective. In the randomized trial, the experimental combination had little effect on response rate. A clear improvement was seen in Ki67, but this biomarker has limited relevance for patients. Thus, although the trial was relatively large it would have no impact on clinical practice.

The analyses for randomized trials designed to directly compare the experimental and control groups are identical to those used in Phase III trials, but with the level of statistical significance specified in the sample size calculation. For any trial design, separate exploratory subgroup analyses based on biomarkers can be performed to determine whether some patients benefit more than others; these markers might then be included in the design of further trials (as a stratification factor or eligibility criterion).

Efficacy and safety events are often analyzed on an ITT basis (see Figure 1.3), although more flexibility is allowed in Phase II studies. For example, the primary analysis may be based only on participants who started the trial treatment, received a sufficient amount of treatment or completed the treatment as specified in the protocol (per protocol analysis). Such analyses maximize the chance of seeing benefit with the experimental treatment and can give a better idea of its direct effect on efficacy and harms.

Conduct

Phase II studies are often conducted in a similar manner as Phase I trials (see Chapter 2), though some processes such as safety and data monitoring might be less intensive.

Effect of bias. Many Phase II studies are conducted by experienced health professionals in specialist centers. Therefore, the observed treatment benefit might be overestimated (particularly if the trial is not blinded) because some of the benefit could be due to the high standard of care in these centers. The results should also be interpreted in the context of the characteristics of the patients who were recruited

because they may, for example, have had a poorer (or better) prognosis than originally anticipated, leading to a lower (or higher) response rate or PFS than assumed in the sample size calculation. This sometimes explains why a study does not meet its objectives.

Key points – Phase II trials

- Phase II trials usually provide preliminary evidence on an experimental intervention in a relatively small number of patients. They tend to be conducted in specialist cancer centers.
- Several different designs can be used in Phase II, including single-arm, randomized and multi-arm trials with several experimental therapies or regimens.
- Phase II trials that show large treatment effects, especially single-arm studies, require careful interpretation.
- The decision to conduct a larger trial should be based on primary and secondary efficacy outcomes, as well as safety, adherence and feasibility.

References

1. Brown SR, Gregory WM, Twelves CJ et al. Designing phase II trials in cancer: a systematic review and guidance. *Br J Cancer* 2011; 105:194–9.

2. Lee JJ, Feng L. Randomized phase II designs in cancer clinical trials: current status and future directions. *J Clin Oncol* 2005;23:4450–7.

3. Rubinstein LV, Korn EL, Friedlin B et al. Design issues of randomised phase II trials and a proposal for phase II screening trials. *J Clin Oncol* 2005;23:7199–206.

4. Sadeghi S, Groshen SG, Tsao-Wei DD et al. Phase II California Cancer Consortium trial of gemcitabine-eribulin combination in cisplatin-ineligible patients with metastatic urothelial carcinoma: final report (NCI-9653). *J Clin Oncol* 2019:37:2682–8.

5. Johnston S, Puhalla S, Wheatley D et al. Randomized phase II study evaluating palbociclib in addition to letrozole as neoadjuvant therapy in estrogen receptor-positive early breast cancer: PALLET Trial. *J Clin Oncol* 2019;37:178–89.

6. Cannistra SA. Phase II trials in Journal of Clinical Oncology. *J Clin Oncol* 2009;27:3073–6.

7. Foster JC, Freidlin B, Kunos CA, Korn EL. Single-arm phase II trials of combination therapies: a review of the CTEP experience 2008–2017. *J Natl Cancer Inst* 2020;112:128–35.

8. Freidlin B, McShane LM, Polley MY, Korn EL. Randomized phase II trial designs with biomarkers. *J Clin Oncol* 2012;30:3304–9.

9. Hyman DM, Puzanov I, Subbiah V et al. Vemurafenib in multiple nonmelanoma cancers with BRAF V600 mutations. *N Engl J Med* 2015;373:726–36.

10. Papadimitrakopoulou V, Lee JJ, Wistuba II et al. The BATTLE-2 Study: a biomarker-integrated targeted therapy study in previously treated patients with advanced non-small-cell lung cancer. *J Clin Oncol* 2016;34:3638–47.

4 Phase III trials

A Phase III trial should enable definitive assessment of a new intervention, with the intention of changing clinical or public health practice. Phase I and II trials mostly evaluate pharmaceutical drugs and RT, whereas Phase III trials are used for any type of intervention, including health service delivery aiming to improve the performance of healthcare processes rather than clinical endpoints in individual patients. Well-designed and -conducted Phase III trials provide the highest quality evidence to inform decision-making for treatment and management guidelines, and to support marketing authorization (licensing) and market access.

Participants

The participants should be representative of those for whom the new treatment would be recommended as part of routine care if the trial meets its objectives.[1] Therefore, eligibility is often less restrictive than in Phase I and II trials. Trials in adults with cancer should not have an upper age limit, unless there are clear scientific reasons, although older patients tend to have more comorbidities and may not have adequate PS, which may restrict their recruitment. Women and ethnic minorities tend to be underrepresented.[2]

Setting

Unlike early-phase trials, which are often conducted in specialist (expert) cancer centers, Phase III trials should involve hospitals that are representative of settings in which the intervention will be administered in practice. Most studies are therefore multicenter and, where feasible, international. Conducting a Phase III trial at a single center may limit the generalizability of its results.

Interventions

Specification of the experimental (new) intervention is the same as outlined in previous chapters. In Phase III trials, the experimental treatment must be compared with a control (comparator). Therefore, these trials are only conducted when there is a reasonable expectation of significant clinical benefit, the size of which is influenced by the choice of comparator. Box 4.1 summarizes key features of a comparator. Deciding on a relevant comparator can be challenging, particularly when the SOC differs between geographical regions and different expectations of regulators and HTA agencies need to be met. Using an inappropriate comparator (for example, outdated treatments or a lower dose than recommended) can lead to overestimation of the benefits of a new treatment and an overly favorable cost-effectiveness assessment.

Blinding

Where feasible and ethical, Phase III trials are double blinded (using placebo), such that neither the patient nor the research team is aware of the assigned interventions until the end of the trial. Knowing

BOX 4.1

Features of a control (comparator) intervention for Phase III trials

- A current SOC (using same dose and frequency of administration as per practice)
- Can be a specific treatment mandated in the protocol, or clinician choice from two or more recommended therapies
- Placebo plus either SOC or BSC, when it is unethical to withhold standard therapies
- Placebo alone (where ethical), when there are no established anticancer treatments for that part of the treatment pathway (for example, maintenance therapy or after several lines of treatment)
- Patients will accept the control treatment if they are randomized to receive it
- Health professionals are willing to randomize patients to the control treatment

which trial treatment is received can influence a patient's behavior, which may affect the outcome (placebo effect); it can also influence the research team when assessing the endpoints (assessment bias) or diagnosing recurrence or progression (ascertainment bias). Use of placebo minimizes these biases.

Trial design

Randomization (random allocation) is the most important design feature of Phase III trials. It is analogous to tossing a coin for each patient: 'heads' for the experimental treatment and 'tails' for the control treatment. In practice, a computer algorithm is employed using methods such as simple random allocation, random permuted blocks, stratified randomization, and minimization. The last two methods incorporate prespecified stratification factors in the allocation process. These ensure good balance between the trial arms for these particular factors when they are known to be correlated with outcomes, because a chance imbalance could create confounding.

Stratification factors could be patient or tumor characteristics (for example, age and tumor stage), and center or geographical location if patient management significantly varies between hospitals or countries. A 1:1 allocation ratio is most commonly used, where half the patients receive the experimental treatment and the other half the control treatment. A 2:1 allocation, in which two-thirds of patients receive the experimental treatment, may be preferred to produce more reliable data on the new treatment, particularly AEs.

Trial objectives and designs. Figure 4.1 shows the three categories of objectives. Historically, most Phase III trials were superiority studies because the new treatment was being compared with supportive care or no treatment. However, as more of those new treatments were incorporated into clinical practice as SOC, patient outcomes have improved substantially. In some cases, it is now difficult to find more effective treatments. For example, about 95% of men with testicular cancer survive following standard surgery, chemotherapy and radiotherapy. A new therapy could therefore only nominally improve the survival rate, and this can only be demonstrated using an overly large trial.

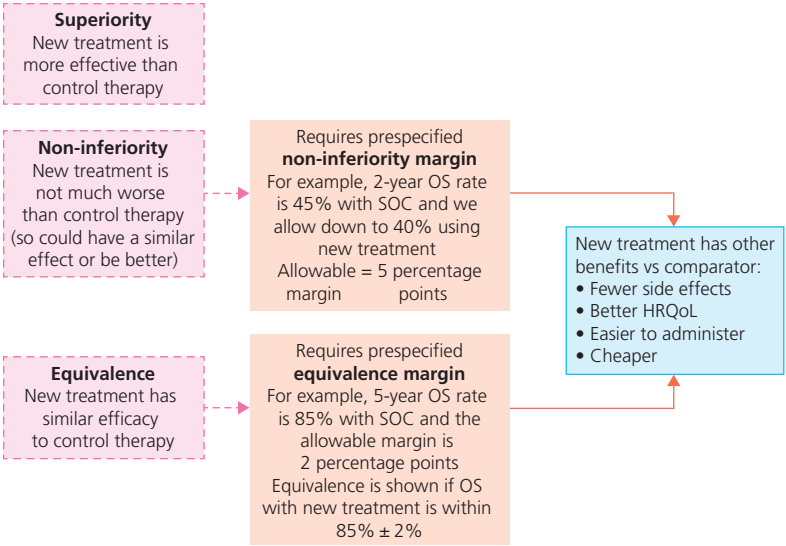


Figure 4.1 Phase III trial objectives. Studies can have more than one objective. For example, a three-arm trial could aim to demonstrate superiority of new drug A over a standard therapy with a different mechanism (for example, pembrolizumab versus cisplatin-based chemotherapy for advanced lung cancer), and non-inferiority of new drug A compared with new drug B in the same class (for example, pembrolizumab versus atezolizumab, both immunotherapies).

Figure 4.2 shows the main types of trial design, which can incorporate adaptive designs.[3] In some cases, the control therapy is a standard treatment and the new treatment represents a different approach, or new class of drug or combination therapy, and is therefore expected to improve efficacy (superiority).

A head-to-head comparison can be used when the control and experimental treatments are of the same type (for example, two immunotherapies) and non-inferiority is the goal because superiority is unlikely to be achieved.[4] In other non-inferiority trials, current SOC may consist of multiple anticancer agents and the goal is to have fewer drugs, lower drug doses or shorter treatment durations to reduce toxicities, without compromising efficacy. The prespecified non-inferiority margin is an essential aspect of the design and interpretation.

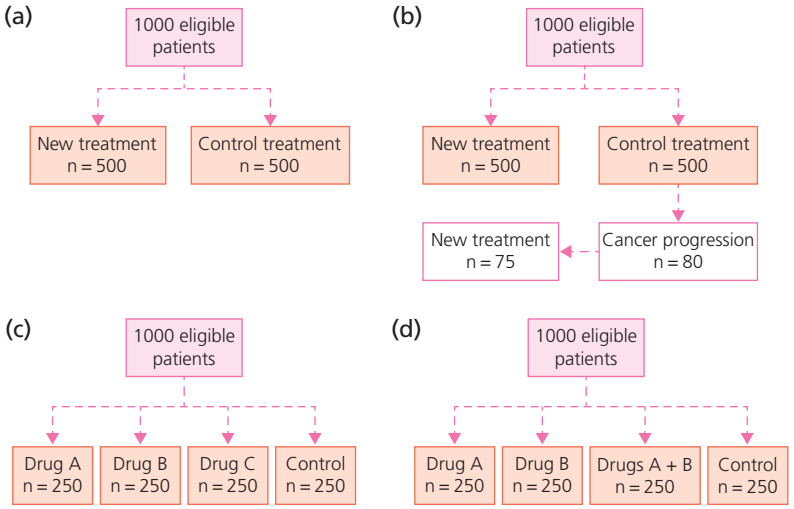


Figure 4.2 Phase III trial designs. (a, b) Parallel studies involving two groups (as in, for example, head-to-head studies). In (a), crossover is not allowed, but in (b), patients whose cancer progresses in the control arm can receive the new treatment. (c) Multi-arm trial. (d) Multi-arm factorial trial. Any of these can also become adaptive designs,[3] which involve reviewing the trial data at several interim analyses and changing the drug regimens, sample size or randomization allocation ratio. There are also multi-arm multi-stage (MAMS) designs, which evaluate several different treatments but with interim efficacy assessments through which futile (ineffective) arms are stopped early, and new treatments groups can be added later. Phase II basket and umbrella trials using biomarkers (such as tumor genetic mutations detected by NGS) to direct experimental therapies can be extended into Phase III by incorporating control arms.

An example of an equivalence trial is the comparison of a generic drug with a marketed formulation, using PK metrics to assess bioequivalence and surrogate biomarkers for clinical effectiveness and safety. The equivalence margin is prespecified.

Factorial trials are an efficient way to evaluate two new interventions at the same time, by determining the effect of each on its own or in combination. This could be two different drugs or an evaluation of both dose and infusion rate of a systemic drug.

Crossover. In some trials, patients in the control arm are allowed to cross over to the experimental therapy if their disease relapses, progresses or recurs, or occasionally after the primary trial results are published. This can make the trial more appealing to patients and recruiting clinicians, knowing that the control patients have the opportunity to receive a (potentially effective) experimental treatment at some point. Crossover can also occur when those in the experimental group have their treatment switched to the control treatment, but this is less common.

However, crossover can make OS uninterpretable if many control patients receive the experimental drug (see Table 1.5). An example is crizotinib for treating anaplastic lymphoma kinase (ALK)-positive advanced non-small-cell lung cancer:[5] in the pivotal trial, 87% of patients randomized to standard chemotherapy (controls) later crossed over to receive crizotinib, yielding no observable impact on OS (HR 1.02). The HTA agency in Germany commented that this looked like a single-arm trial and was unable to make conclusions about OS.

Trial outcome measures

Phase III trials are designed to collect data relating to efficacy, safety (AEs), adherence to the trial treatments and any other anticancer treatment, PROs and health resource utilization to support pharmacoeconomics.[6,7] Primary outcome measures should have some of the following features:

- clinically relevant to patients and healthcare professionals
- clearly defined and easily and reliably quantified
- a measure of wellbeing or HRQoL
- potential societal impact, including for healthcare delivery.

Commonly used time-to-event endpoints are shown in Table 1.5. Tumor response is unlikely to be accepted as a primary outcome in most Phase III studies of solid tumors, though complete remission may possibly be used for some hematologic malignancies. Other efficacy measures may be specific to the type of intervention being evaluated and disease, such as the complete resection rate after neoadjuvant chemotherapy for early-stage cancer, where the primary purpose of treatment is to shrink the tumor to improve the likelihood of successful surgical resection; also see Table 5.3 for other types of endpoints (used in RT trials).

Overall survival is regarded as the most objective and relevant primary endpoint in Phase III oncology trials. It requires only the date of death (not the cause), but it has limitations and is not often used for early-stage (curable) cancers.

Surrogate endpoints are commonly used in early-phase trials, but their role as a primary endpoint in Phase III trials is less clear (see page 17).[8] Trials that show substantial improvement in PFS but little or no impact on OS represent a challenge for decision-makers in determining the value of a treatment.

Regulatory agencies can approve a new treatment on the basis of PFS alone, although the license might be withdrawn later on if subsequent trials show no OS benefit. However, benefits on PFS alone that are accompanied by reductions in AEs or improvements in HRQoL may be viewed favorably by decision-makers.[9] For uncommon tumor types (or subtypes) with poor prognosis or clear unmet need, a large benefit in PFS could lead to marketing authorization and possibly reimbursement and market access approval, though HTA agencies and payers will scrutinize the PFS data carefully.

Trials may however include coprimary endpoints (OS and PFS), ensuring that the study is powered to show benefits in either; decision-makers can then consider the merits of each endpoint when the results are available. Also, the final PFS data are likely to be ready before OS enabling decisions to be made earlier, such as regulatory submissions for a license.

Patient-reported outcomes such as HRQoL, which include individual symptoms (see Box 1.1), and measures of patient experience, contribute to better care in addition to clinical efficacy. It is hoped that a new treatment will help to at least maintain HRQoL by stabilizing or reducing cancer-related symptoms, particularly when added to standard treatment. Occasionally, new drugs used as monotherapy have led to substantial improvements in HRQoL when compared with standard toxic chemotherapy combinations. Regulators and HTA agencies are interested in the quality of survival and may not approve a drug that improves survival only moderately but substantially compromises HRQoL.

Health resource utilization data are often collected, such as number of days in hospital and visits to the family physician, costs of treating AEs and costs incurred by patients when attending clinic for treatments and follow-up investigations. These are used in cost-effectiveness analyses.

Assessments of disease are done by local clinicians and pathologists (for example, RECIST for solid tumors or disease-specific criteria for hematologic malignancies). Blinded independent central review (BICR) of diagnostic scans and pathology specimens, where the central reviewer (for example, radiographer) does not know the treatment allocation or efficacy outcomes, minimizes assessment bias that might occur among local reviewers. BICR can mitigate criticism by regulatory and HTA agencies who might otherwise question the reliability of trial results, particularly in open-label studies. The increasing use of artificial intelligence should reduce this bias and make the review process cheaper and easier.

Study size

Phase III trials typically involve several hundred or, occasionally, a few thousand patients. The fundamental principles of estimating sample size are shown in Figure 1.2. Two key parameters are the expected treatment effects in the experimental and comparator arms; the larger the difference, the smaller the trial required. Non-inferiority and equivalence trials require specification of an allowable margin (see Figure 4.1), and they involve more patients than a superiority study, mainly because the allowable margin should be smaller than what would be considered a clinically important improvement in efficacy in a superiority study. The margin must be acceptable to patients, clinicians and decision-makers and may be derived from discussion with patient representatives and the consensus of several clinicians.

Analysis and interpretation of results

There is a vast literature on how to analyze and interpret Phase III trials.[10,11] All trial endpoints can be divided into three categories of data and the categorization determines the method of sample size calculation, statistical analysis and appropriate type of effect size.

- Continuous data involve taking measurements from patients, such as tumor size (mm), HRQoL score and biomarker level.
- Binary/categorical data are counts of people in mutually exclusive groups, such as dead/alive or recurrence/no recurrence (both ignoring time), and tumor response (CR, PR, SD and progressive disease).
- Time-to-event data measure the time it takes for a predefined event to occur: for example, OS, PFS and DFS. OS is the same as dead/alive, but it incorporates the length of time before a patient died.

Effect size is a single number produced by comparing a particular endpoint between two groups (mostly by taking the difference or their ratio). It provides a way to interpret and communicate the efficacy results of a new treatment (Table 4.1). Effect sizes can also be used for AEs, adherence and HRQoL. Several features need to be considered when examining effect sizes (Figure 4.3).

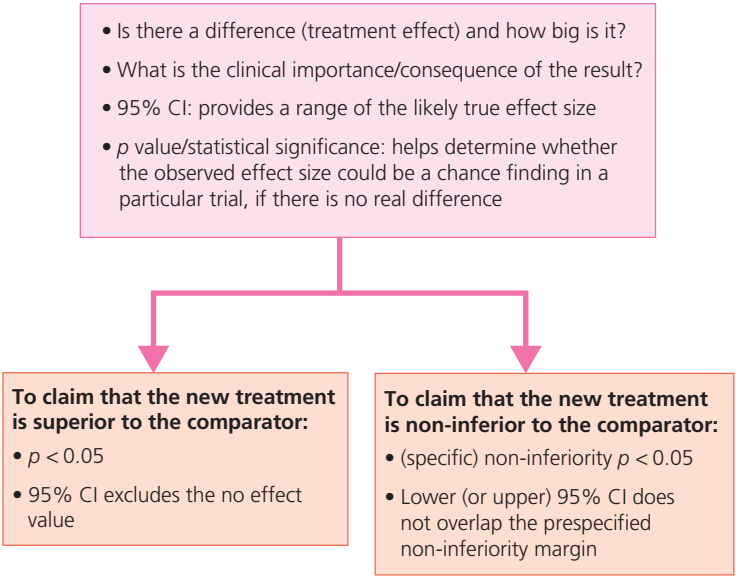


Figure 4.3 General approach to interpreting effect sizes. The 'no effect' value is the value of the effect size when the new treatment has the same effect as the control (0 for differences; 1 for HR, odds ratio and relative risk). The concepts can be used for any endpoint. CI, confidence interval.

TABLE 4.1

Commonly used effect sizes and statistical analyses

Category of data	Effect size	Statistical analyses that produce an effect size, CI and *p* value
Continuous data (taking measurements from patients)	• Difference between two means • Difference between two medians • Regression coefficient (slope)	• *t* test • Mann–Whitney test • Linear regression
Binary/categorical data (counting people in separate groups)	• Relative risk (risk ratio)* • Rate ratio • Odds ratio • Absolute risk difference • Number needed to treat	• Chi-squared (χ^2) test • Logistic regression
Time-to-event data	• Hazard ratio* • Difference between two medians At a specific time point: • Absolute risk difference • Number needed to treat	• Kaplan–Meier curves • Logrank test • Cox proportional hazards regression

*Unlike HR, relative risk ignores the time taken for each patient to reach an event such as death: the number of events at the end of a trial is expressed as a percentage of the number of patients randomized in each group (the ratio of these percentages is the relative risk). Because relative risk ignores time, it can be less sensitive than HR. Authors sometimes use the term 'relative risk' in a publication whereas the methods section indicates that they have actually provided the HR. CI, confidence interval.

Figure 4.4 presents an example of a Phase III superiority trial,[12] from which three effect sizes can be obtained using the same time-to-event endpoint. The interpretation of these results is shown in Box 4.2. Table 4.2 provides a commentary on each effect size. The interpretation of benefit associated with a new therapy depends on the particular effect size used. Different types of effect sizes often lead to the same conclusion, particularly when the treatment effect is very large or there is no benefit. However, an HR can sometimes give the impression of a large benefit (because it only measures relative effects), whereas the difference in median OS (or PFS) or difference in event rates at specific timepoints indicate only small or moderate benefits (absolute effects).

HR is appropriate when the assumption of proportional hazards is met (see footnote to Table 4.2). But when the Kaplan–Meier curves clearly cross over or they have unusual shapes (for example, large benefit early on then small/no benefit later on, or vice versa) a single HR cannot accurately reflect this. It is better to compare the event rates at specific time points (landmark analysis), or the medians. It is also possible to compare the area under the curves (called restricted mean survival time [RMST]: a measure of the average duration of survival over the trial follow-up if using OS, sometimes referred to as life expectancy, or average PFS using PFS).[13]

Interpretation of a non-inferiority trial is focused on the confidence interval of the effect size, rather than the point estimate itself (see Figure 4.3). An example of a non-inferiority trial is provided in Chapter 5 (see Figure 5.1).

Trials of cancers with good prognoses. Figure 4.5 shows the results of a trial of first-line treatment of follicular lymphoma.[14] Patient outcomes are so good that the median PFS was not reached for either treatment, including the SOC. OS is high and similar between the two treatments, partly because the data are immature. This example illustrates the challenge of showing improvements in OS in cancers with a good prognosis. For some cancers, such as indolent hematologic malignancies, time to next anticancer treatment (TTNT) could be considered a clinically relevant outcome because it directly reflects a change in patient management. In this trial, 16% of the obinutuzumab group had started another treatment, compared with 23% of the rituximab

Patients: NSCLC that has progressed after previous treatment, 1034 patients randomized; 442 with ≥ 50% PD-L1 expression

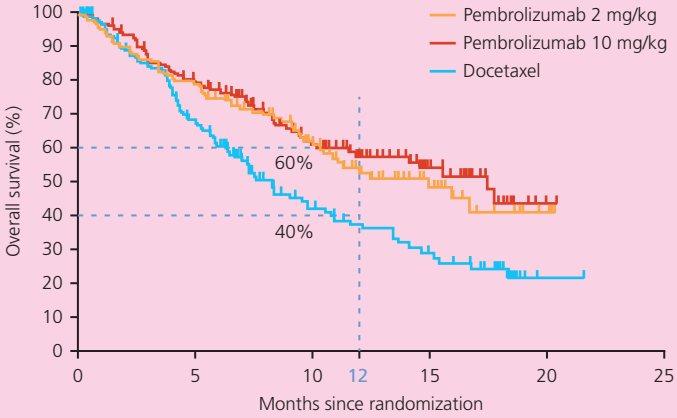
Interventions: Pembrolizumab, 2 and 10 mg/kg, vs docetaxel (comparator)

Primary endpoint: OS in patients with ≥ 50% PD-L1 expression

Design: Superiority trial (target HR 0.55)

Key results:

Treatment	Median OS (months)	HR Pem vs docetaxel (95% CI)	*p* value
Pembrolizumab 2 mg/kg	14.9	0.54 (0.38, 0.77)	0.0002
Pembrolizumab 10 mg/kg	17.3	0.50 (0.36, 0.70)	< 0.0001
Docetaxel	8.2		



Number at risk

Pembrolizumab 2 mg/kg	139	110	51	20	3	0
Pembrolizumab 10 mg/kg	151	115	60	25	1	0
Docetaxel	152	90	38	19	1	0

The number at risk decreases over time as patients have the event (death in this instance) or are censored

- Large separation between the curves should indicate large treatment effects
- Each upward dash on the plots represents a censored patient (i.e. alive at that time point but no information afterward)
- There is little/no difference in OS between the three groups up to 3 months. The higher dose of pembrolizumab seems to have a slightly greater benefit than the lower dose after 12 months (to be considered alongside toxicity)
- Survival rates at specific time points can be found by drawing a vertical line (for example, at 12 months)
- Very few patients at risk from 20 months, so any treatment effect here is unreliable

Figure 4.4 KEYNOTE-010, an example of a Phase III trial. NSCLC, non-small-cell lung cancer; Pem, pembrolizumab; vs, versus. The graph is reproduced, with permission, from Herbst et al. 2016.[12]

BOX 4.2

Interpretation of the three time-to-event effect sizes from the trial in Figure 4.4

Hazard ratio

- The risk of dying at any time (having survived to that time) was reduced by 50% (HR 0.50) with 10 mg/kg pembrolizumab versus docetaxel, and by 46% (HR 0.54) with pembrolizumab, 2 mg/kg; both exceed the target HR of 0.60. This is a clinically large benefit
- 95% CI for 10 mg/kg pembrolizumab: the true treatment effect (true HR) is likely to be somewhere between 0.36 and 0.70 (if 100 identical trials were conducted, the intervals for 95 should contain the true effect)
- *p* value <0.0001 for 10 mg/kg pembrolizumab: means that if it is assumed pembrolizumab truly has the same effect as docetaxel (true HR 1.0), then an effect as large as HR 0.50 (or more extreme HR ≤ 0.50, or in the opposite direction, which is HR ≥ 2 [1/0.5]) could be seen by chance alone but only in fewer than one trial among 10 000 with the same design. Therefore, the observed HR is highly unlikely to be due to chance and so likely to reflect a real treatment benefit

Difference in median OS

- Median OS improved from 8.2 to 17.3 months with 10 mg/kg pembrolizumab versus docetaxel (loose interpretation: patients lived, on average, 9.1 months longer). This is a clinically moderate/large benefit

Absolute risk difference at a time point

- At 12 months, the absolute risk difference is 20 percentage points (60–40%): among 100 patients treated with 10 mg/kg pembrolizumab an *additional* 20 could be alive, compared with 100 given docetaxel. This is a clinically moderate/large benefit

The data are from KEYNOTE-010.[12] CI, confidence interval.

TABLE 4.2

Three types of effect sizes for time-to-event data*

Effect size	Interpretation	Comments
Hazard ratio 0.50	Risk of dying is decreased by 50% (at any single time point, having survived to that time)	• Measure of relative effect • Uses the whole Kaplan–Meier curve • Assumes proportional hazards† • Efficacy assessed using risk
Median OS (17.3 vs 8.2 months)	Median OS increased by 9.1 months with pembrolizumab	• Measure of absolute effect • Uses only one point on each curve (can be influenced by chance) • Easy for patients and clinicians to understand • Efficacy assessed using time
Survival rate at 12 months 60% vs 40% (called landmark analysis)	20 *more* patients alive at 12 months among 100 who received pembrolizumab	• Measure of absolute effect • Uses only one time point on each curve (can be influenced by chance) • All patients should be followed up to the time point, unless they had the event before • Useful for quantifying a plateau in the tail of the Kaplan–Meier curves (long-term benefit) • Indicates 'impact' in a group of treated patients • Efficacy assessed using risk

*Refer to the example in Box 4.2. The three effect sizes indicate different aspects of the efficacy of a new therapy.

†The risk of an event (death in this example) at a single time point is called a hazard; the ratio of two hazards should be the same at all time points (except at the very start).

vs, versus.

group.[14] Other useful endpoints for this cancer type include negative minimum residual disease.

Statistical significance. A *p* value essentially addresses the question: could the observed effect size be a chance (spurious) finding in this trial, assuming that the new treatment is truly no better than the comparator? The answer to this question is always 'yes', but *p* values are a way of quantifying this. All *p* values lie between 1 (definitely no effect) and 0 (definitely an effect). Large treatment effects move *p* values toward 0, as does having many patients or events. The smaller the *p* value, the stronger the evidence of a real treatment effect (see example in Box 4.2).

A cut-off of $p = 5\%$ (0.05) is used to make decisions: $p < 0.05$ indicates that the effect size is statistically significant (likely to represent a real effect, though this is not guaranteed), whereas $p \geq 0.05$ indicates that a difference is not statistically significant. This cut-off has no scientific basis but has become ingrained in practice.

Statistical significance must be distinguished from *clinical importance*, to avoid a common misinterpretation of *p* values. Lack of statistical significance ($p \geq 0.05$) does not mean that the new treatment has no effect, only that there is insufficient evidence of an effect. For example, a 45% reduction in deaths with $p = 0.12$ means that the effect could be, but is not necessarily, a chance finding. But it is incorrect to interpret this as having no effect at all only because the *p* value is above 0.05.[15,16] However, clinical importance focuses attention on the effect size itself (and its confidence interval), not the *p* value, so both need to be interpreted together. A *p* value just above 0.05 (for example, 0.07) associated with a clear treatment benefit is usually due to insufficient patients or events; the latter could be avoided by having longer follow-up.

Intention-to-treat analysis is the standard because it retains the balance in patient characteristics achieved by randomization (see Figure 1.3). ITT analysis is always used for superiority trials.

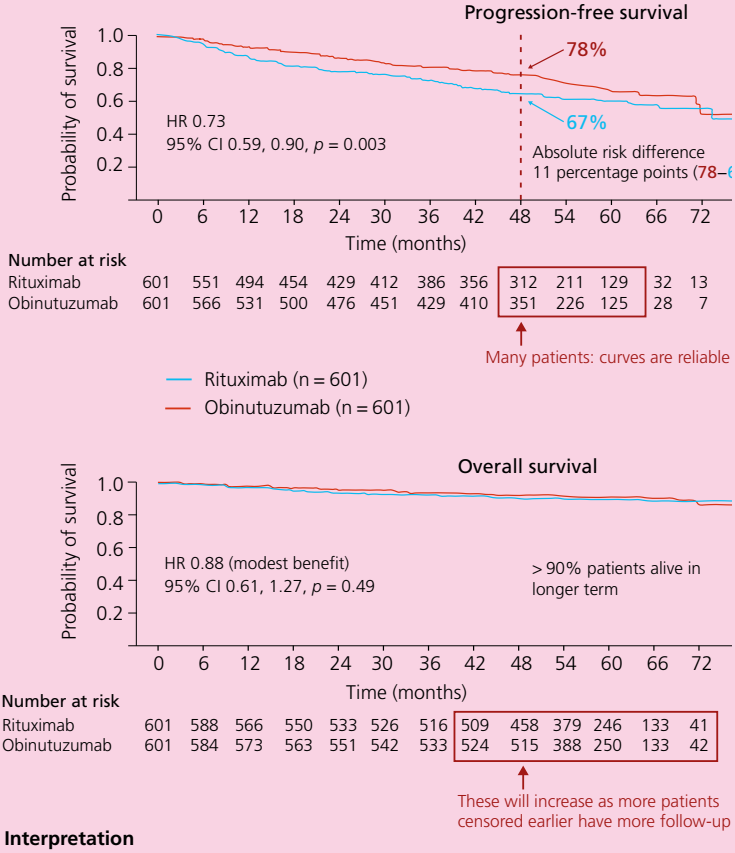
Per protocol analysis can be used in addition to ITT: only patients who started and adhered to the allocated trial treatment (as specified in the protocol) are included in the analysis. Both ITT and per

Patients: Follicular lymphoma; no prior treatment

Interventions: Obinutuzumab (experimental) vs rituximab (comparator) each with standard chemotherapy

Outcome: PFS

Design: Superiority trial (target HR 0.74)
1202 patients randomized



Interpretation

- PFS: close to the expected effect; benefit for obinutuzumab seen at 4 years and beyond
- OS: most patients are alive, even at 5 years, so unlikely to ever show a clear OS benefit

Figure 4.5 The GALLIUM study, an example of a trial in a tumor type with good prognosis. CI, confidence interval; vs, versus. The graphs are reproduced, with permission, from Townsend et al. 2018.[14]

protocol analyses should be performed for non-inferiority trials, although the latter is often preferred because it is less affected by non-adherence to the trial interventions.

Analyses of AEs could include only patients who started the trial treatments, referred to as the safety dataset or population.

Subgroup analyses

This involves analyzing patients within separate groups according to a factor (for example, sex) and examining the effect size in each subgroup (males and females in the example). The four possible outcomes of subgroup analyses are illustrated in Table 4.3. Situations in which the treatment is beneficial in one group of patients but has no effect or is harmful (lower efficacy than the comparator treatment) in another group have potentially important consequences if the recommendation is that the new treatment is offered only to some patients and not others.

The following criteria can ensure the reliability of a subgroup claim.[17]

- The confidence interval (CI) for at least one subgroup does not overlap with the *overall treatment effect* based on all trial patients. A common mistake is to compare each interval with the no-effect value (for example, 1 for HR or 0 for absolute risk difference).
- The test for interaction is significant ($p < 0.05$, ideally < 0.01), meaning that the effect size in one subgroup is significantly different from that in the other.
- The differential effect of the treatment between the groups is biologically plausible.
- The finding is corroborated by evidence from independent studies.

Spurious subgroup effects may be seen when analyzing many factors, so any claim should be viewed with caution, particularly when the main trial objective based on all patients has not been achieved.

HRQoL can be summarized in different ways (Figure 4.6). In advanced cancer, HRQoL is expected to deteriorate with worsening of cancer-related symptoms as the disease progresses. It is therefore important to determine whether the new treatment can slow or delay the onset of these symptoms, or at least have no effect on them, and thus maintain (or even improve) HRQoL. Some drugs have serious AEs that compromise HRQoL.

TABLE 4.3

Possible conclusions from subgroup analyses, illustrated using sex as the factor

HR, new versus control treatment (for example, using OS or PFS)		
Males	Females	Conclusion
0.75	0.73	• Size of treatment benefit is similar between the two subgroups – no subgroup effect • Provides reassurance that the treatment can be recommended for all patients
0.45	0.75	• Treatment is effective in all patients but the benefit is greater in males • All patients should still be offered the new treatment, but a cost-effectiveness evaluation might indicate it is only worth doing so in males
0.65	0.95	• Treatment is beneficial in males but has the same effect as the control therapy in females • *Perhaps* only males should be offered the new treatment
0.65	1.50	• Treatment is beneficial in males but worse than control (i.e. harmful) in females • *Perhaps* only males should be offered the new treatment

A statistical 'test for interaction' compares the HRs between the two groups (for example, 0.75 versus 0.73). $p < 0.05$ indicates that the HRs are numerically different, suggesting a differential treatment effect, while $p \geq 0.05$ does not provide sufficient evidence for a subgroup effect.

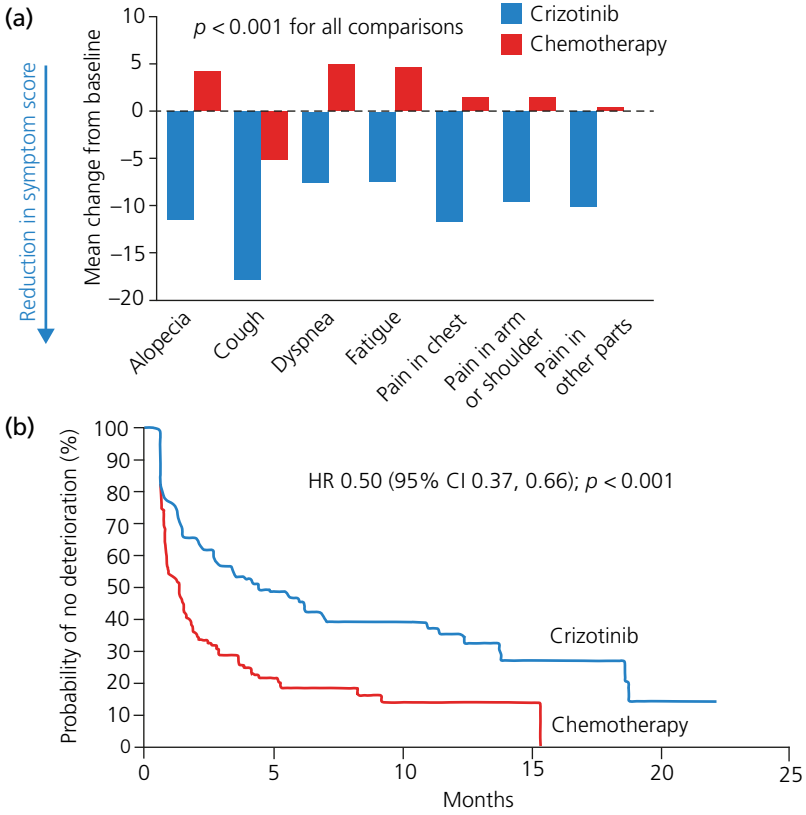


Figure 4.6 Two fundamentally different ways of showing HRQoL data using a trial of crizotinib versus standard chemotherapy in patients with advanced ALK-positive non-small-cell lung cancer.[5] (a) The bar chart uses the actual scores for each patient (summarized as mean values). All of the following need to be understood to interpret these results: (1) the scale for each symptom measure (0–100 here); (2) whether a decreased score from baseline reflects improvement or worsening of symptoms (improvement here); and (3) what defines a clinically important change in score (5–10 points for this measure). Several symptoms improved noticeably with crizotinib. (b) The scores have been converted to time-to-event data for each patient (i.e. risk); the Kaplan–Meier curve shows the time to deterioration in any of three symptoms: cough, dyspnea and chest pain. A clinically important size of deterioration must be defined (here, 10-point increase from baseline). HR indicates whether the risk of deterioration (by 10 points) is increased or decreased versus chemotherapy. Here the risk is halved: a large improvement. The graphs are reproduced, with permission, from Shaw et al. 2013.[5]

Other outcome measures. The importance of collecting and reporting data on AEs[18] and adherence to trial treatments has been summarized in Chapter 1. A clear difference in the incidence of a particular AE might be tested for statistical significance, but biological plausibility and impact are more important.

Trials using biomarkers

Biomarkers are an increasingly important aspect of Phase III trials.[19] For example, biomarker analyses could explain why only some patients benefit from treatment. Occasionally, funding from public sector or charitable organizations is contingent on translational research being part of the protocol.

Collection of tumor samples (and sometimes tissue from unaffected areas) and bone marrow aspirate is invasive and therefore these samples are often only collected at baseline or during surgery. Imaging scans may also be requested for the measurement of novel markers.

Biomarkers (or any other patient or tumor factor) can be evaluated as either prognostic or predictive markers (Figure 4.7). There is often interest in finding an effective predictive marker that can direct anticancer treatment (personalized or stratified medicine). This is best evaluated using subgroup analyses in trials of unselected patients, ideally when the choice of biomarker has been prespecified in the protocol.

Examples of successful identification of predictive markers are panitumumab for *KRAS* wild-type metastatic colorectal cancer and pembrolizumab for advanced non-small-cell lung tumors that express a threshold level of PD-L1.

Regulatory and HTA agencies tend to be critical of analyses involving retrospectively identified biomarkers.

Overall interpretation of benefit

The evaluation of the benefit of an experimental drug or combination will be based on the effect sizes from primary and key secondary endpoints, and will also take into account AEs and impact on HRQoL and other PROs, and the quality of survival.

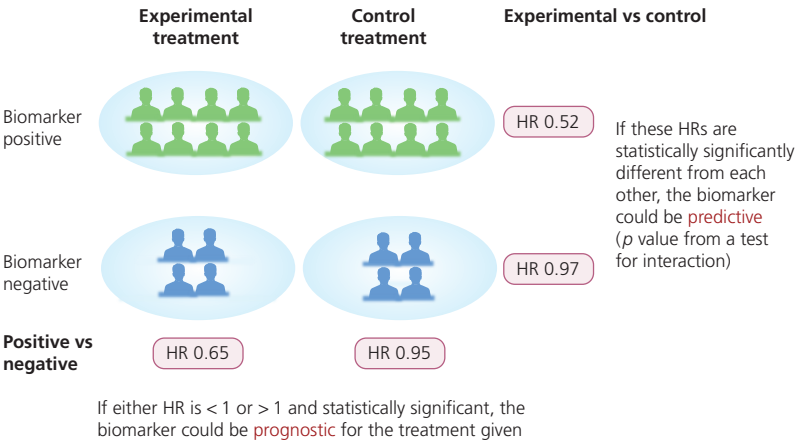


Figure 4.7 Illustration of the difference between a prognostic and predictive biomarker (outcome could be OS). Patient and disease characteristics are similar between the experimental and control groups (due to randomization) but possibly different when comparing biomarker-positive and -negative patients (multivariable regression analyses can allow for this). Prognostic marker: compares patients who are biomarker positive and negative among those given the same treatment. In the figure, the marker is only prognostic in the experimental group (biomarker-positive patients are less likely to die than those who are biomarker negative; HR 0.65). Predictive marker: when the effect of the experimental treatment differs between patients who are biomarker positive and negative. In the figure, 0.52 is different from 0.97 (new treatment is beneficial for marker-positive but not marker-negative patients). Similar HRs would mean that the effect of the new treatment does not depend on the biomarker (i.e. it is not a predictive marker).

Key points – Phase III trials

- Phase III trials are the most robust approach for evaluating a new intervention against SOC or a placebo. Phase III trials frequently evaluate new pharmaceutical agents but can be used to assess a new surgical intervention, RT technique or health service delivery policy.
- Randomization is central to the design of Phase III trials. The trial must be of sufficient size to demonstrate a significant improvement in clinically relevant endpoints if it is to influence clinical practice.
- Trials must include the patient population that is intended to benefit from the new intervention.
- A variety of primary and secondary endpoints can be used; the choice depends on the type of cancer and whether it is at an early or late stage (i.e. good or poor prognosis).
- Trial analyses must be outlined in the protocol and accurately reflected in the final publication of results.
- When interpreting the results of trials, the clinical value of the observed treatment effect must be considered alongside statistical significance. Subgroup analyses require great care.

References

1. Kim ES, Bruinooge SS, Roberts S et al. Broadening eligibility criteria to make clinical trials more representative: American Society of Clinical Oncology and Friends of Cancer Research Joint Research Statement. *J Clin Oncol* 2017;35:3737–44.

2. Geller S, Koch A, Roesch P et al. The more things change, the more they stay the same: a study to evaluate compliance with inclusion and assessment of women and minorities in randomized controlled trials. *Acad Med* 2018;93:630–5.

3. Bhatt DL, Mehta C. Adaptive designs for clinical trials. *N Engl J Med* 2016;375:65–74.

4. Mauri L, D'Agostino RB Sr. Noninferiority trials. *N Engl J Med* 2018;378:304–5.

5. Shaw AT, Kim DW, Nakagawa K et al. Crizotinib versus chemotherapy in advanced ALK-positive lung cancer. *N Engl J Med* 2013;368:2385–94.

6. US Department of Health and Human Services et al. *Clinical Trial Endpoints for the Approval of Cancer Drugs and Biologics: Guidance for Industry*, 2018. www.fda.gov/media/71195/download, last accessed 2 June 2020.

7. Hackshaw AK. *A Concise Guide to Clinical Trials*. Chichester: Wiley, BMJ Books, 2009.

8. Korn EL, Freidlin B. Surrogate and intermediate endpoints in randomized trials: what's the goal? *Clin Cancer Res* 2018;24:2239–40.

9. Cherny NI, Dafni U, Bogaerts J et al. ESMO-Magnitude of Clinical Benefit Scale version 1.1. *Ann Oncol* 2017;28:2340–66.

10. Pocock SJ, Stone GW. The primary outcome is positive – is that good enough? *N Engl J Med* 2016;375:971–9.

11. Pocock SJ, Stone GW. The primary outcome fails – what next? *N Engl J Med* 2016;375:861–70.

12. Herbst RS, Baas P, Kim DW et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* 2016; 387:1540–50.

13. Dehbi HM, Royston P, Hackshaw A. Life expectancy difference and life expectancy ratio: two measures of treatment effects in randomised trials with non-proportional hazards. *BMJ* 2017;357:j2250.

14. Townsend W, Buske C, Cartron G et al. Obinutuzumab-based immunochemotherapy prolongs progression-free survival and time to next anti-lymphoma treatment in patients with previously untreated follicular lymphoma: four-year results from the phase III GALLIUM study. *Blood* 2018:132(suppl 1):1597.

15. Hackshaw A, Kirkwood A. Research and methods: interpreting and reporting clinical trials with results of borderline statistical significance. *BMJ* 2011;343:d3340.

16. Greenland S, Senn SJ, Rothman KJ et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016;31:337–50.

17. Dehbi H, Hackshaw A. Investigating subgroup effects in randomized clinical trials. *J Clin Oncol* 2016;35:253–4.

18. Lineberry N, Berlin JA, Mansi B, Glasser S. Recommendations to improve adverse event reporting in clinical trial publications: a joint pharmaceutical industry/journal editor perspective. *BMJ* 2016; 355:i5078.

19. Freidlin B, Korn EL. Biomarker enrichment strategies: matching trial design to biomarker credentials. *Nat Rev Clin Oncol* 2014;11:81–90.

5 Trials of non-drug interventions

There are multiple clinical trials that assess surgical procedures, RT and cell and gene therapies. Other trials evaluate interventions impacting behavioral or lifestyle changes to improve symptoms and HRQoL. The fundamental principles of clinical trials apply to these non-drug interventions, but each type of intervention has unique characteristics and challenges. Generally, these include lack of blinding, the need for good-quality assurance of how the intervention is administered, and issues over genuine equipoise and randomization.

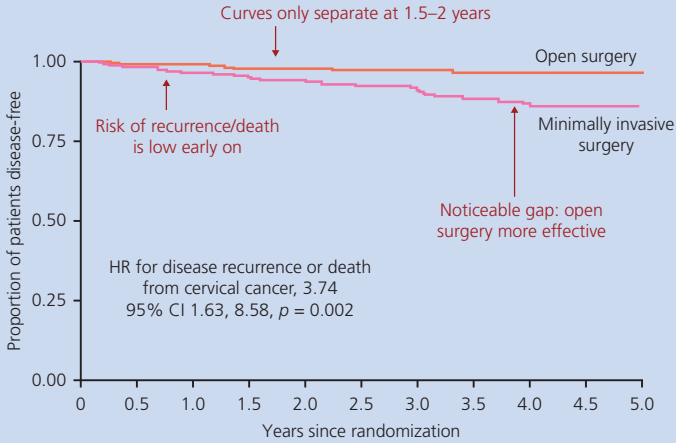
Many clinical trials of new pharmaceutical agents are sponsored by the manufacturers, who have significant resources for design and conduct, which also accommodate the regulatory and market access requirements in a multi-institutional setting. In contrast, many trials of surgical procedures and RT are conducted by small groups of specialist surgeons or radiation oncologists, sometimes in only a few institutions.

Surgery

Surgery (complete tumor resection) often has curative intent, but other surgical procedures such as debulking of large tumors, removal of blockages caused by tumors and removal of small metastatic lesions from the liver or lungs can also be used to improve symptoms or delay disease progression. Surgery may be the only treatment needed for early-stage disease. However, high-quality published surgical trials are lacking. In the past, an estimated one in five surgical trials (all disorders) was discontinued prematurely because of lack of accrual, and the results of one-third were not published.[1] There are challenges to developing new surgical techniques and evaluating them in randomized trials.[2,3] Figure 5.1 and Table 5.1 describe examples of two surgical trials with different objectives (non-inferiority and superiority), which are covered in this chapter.

Well-conducted surgical trials can change practice, occasionally in unexpected ways. For example, the international Laparoscopic

Patients:	Newly diagnosed cervical cancer, stage IA–IB1; PS 0–1 631 patients randomized
Interventions:	Radical hysterectomy using minimally invasive surgery (MIS; laparoscopy or robot-assisted) vs open radical surgery (ORS)
Outcome:	DFS at 4.5 years
Design:	Non-inferiority (allowable margin –7.2 percentage points)
Compliance with assigned surgery:	MIS vs ORS: 91% vs 88%
Main result:	DFS at 4.5 years: MIS vs ORS: 86% vs 96.5% Risk difference: –10.6% (95% CI –16.4, –4.7)



Number at risk											
Open surgery	312	280	236	187	163	144	134	123	104	90	7
Minimally invasive surgery	319	292	244	192	167	155	142	121	102	80	5

Other key result:	Number of days in hospital: 3 vs 5
Interpretation:	About 11 fewer patients (per 100 given MIS) were alive and cancer-free after MIS vs ORS at 4.5 years

This difference is clinically significant; the CI includes the allowable margin (–7.2 percentage points) so non-inferiority is not achieved. It is also statistically significant (CI excludes 0) in favor of ORS.

Figure 5.1 Example of a surgical trial that did not meet its intended objective of non-inferiority (the Laparoscopic Approach to Cervical Cancer trial). MIS, minimally invasive surgery; ORS, open radical surgery; vs, versus. The graph is reproduced, with permission, from Ramirez et al. 2018.[4]

TABLE 5.1

Example of a surgical trial that met its intended objective of superiority

Patients:	Breast cancer, stage 0–III; partial mastectomy planned 235 patients randomized
Interventions:	Tumor resection with additional tissue removed around the margins (shave) vs resection without shaving (margin defined as per usual practice)
Outcome:	Positive margin on the removed tissue (tumor touches the edge of the specimen), by pathology assessment
Design:	Superiority trial: expected positive margin rate 15% (shave) vs 30% (no shave)
Setting:	Single center (4 surgeons)
Compliance with assigned surgery:	100% for both interventions
Results:	Positive margin rate: 19% vs 34% ($p=0.01$) Second operation for margin clearance: 10% vs 21% ($p=0.02$)
Interpretation:	The use of cavity shave margins significantly reduces the number of women with positive margins, which also reduces the need for a second procedure

From Chagpar et al. 2015.[6]
vs, versus.

Approach to Cervical Cancer (LACC) study in the treatment of early-stage cervical cancer found a higher recurrence rate with minimally invasive surgery (hysterectomy by laparoscopic or robotic surgery) than with open radical surgery (laparotomy; see Figure 5.1).[4] This finding led to worldwide reconsideration of the use of

laparoscopic and robotic surgery, which had become SOC for this patient group in many centers. This example illustrates how a new surgical approach had been adopted into practice in the absence of reliable evidence from randomized studies, because the benefits (shorter hospital stay and less surgical morbidity) were considered to outweigh a presumed minimal impact on efficacy (recurrence and deaths).

Recruitment. Many eligible patients decline to be randomized in surgical studies; they may expect the surgeon to know which surgical procedure is best, or the surgeon has a preference that inadvertently influences the patient's decision. Only surgeons with genuine equipoise should participate in a randomized trial. Even when many surgeons formally sign up to a study, actual recruitment can be low if they find reasons for individual patients to receive one procedure in preference to another.

In the example in Figure 5.1, it took 9 years to recruit 631 patients, even with 33 international sites. Uptake can be improved by close collaboration with patient representatives to develop well-written information about the trial, accompanied in some cases by a short video for potential patients. There could also be support for surgeons when describing the trial, including the concept and importance of randomization.[5]

Interventions. Whereas pharmaceutical agents can be manufactured in a standardized way, surgical procedures combine the surgeon's experience and techniques, choice of surgical instruments and postsurgical care. This makes it difficult to specify exactly how to deliver a new procedure across many centers.

In the example in Figure 5.1, it could be argued that the experimental surgical arm (laparoscopy or robot-assisted surgery) represents two different procedures, even though the efficacy results were similar. Furthermore, techniques such as robotic surgery may only be available in specialized centers with specific surgical expertise. In the example in Table 5.1, the protocol specified where additional breast tissue should be removed within the 'shave' cavity arm (in relation to the tumor) and the surgeons were expected to follow this.[6] However, the volume of tissue could not be standardized in the protocol because it depended on the tumor size and body shape of each patient.

Comparator. The choice of comparator is critical. It can be difficult to recruit to a trial comparing a surgical intervention with no surgery because patients with cancer find a non-intervention arm unacceptable. Possible exceptions are:

- when patients receive other anticancer therapies around the same time, so they perceive the control group as 'less treatment' rather than 'no treatment'
- there are clear benefits to not having surgery (for example, avoiding long-term nerve damage)
- the chance of recurrence is very low, and it is easily treated if it does occur.

Many surgical trials compare procedures that are already used in practice. Therefore, techniques to be evaluated sometimes require a non-inferiority design (Figure 5.1) rather than superiority (Table 5.1).

Managing potential bias. It is impossible to blind the treating surgeon, or the patient in many cases, so in most surgical trials everyone knows which intervention the patient received. This lack of blinding (masking) can lead to various biases, including:

- allocation bias (surgeon decides which intervention to give, or withdraws the patient if they are randomly allocated to the trial arm the surgeon feels is inappropriate)
- performance bias (surgeons change their background behavior or techniques according to which surgical method they are asked to deliver)
- assessment bias (researchers measuring the endpoints are influenced by knowing the intervention allocation).

It is difficult or impossible to develop placebo (sham) surgery in surgical trials. However, bias can be mitigated in several ways by:

- ensuring that the person making the assessments during follow-up is unaware of the treatment allocation and patient outcomes (in Table 5.1, the pathologists who assessed the positive margin status did not know the surgical intervention used for each patient)
- revealing the random allocation as close as possible to the delivery of the surgical procedure (in Table 5.1, instructions about which procedure to follow were provided in sealed envelopes, which were opened only after the partial mastectomy had been performed)

- having a research nurse or other health professional explain the trial to patients and answer queries, as they may be less biased than the surgeon.

Outcome measures. Efficacy endpoints are generally the same as for drug trials. For early-stage cancers, where treatment is intended to be curative, DFS or RFS (or event-free survival) might be the only feasible primary outcome. In Figure 5.1, the DFS Kaplan–Meier curve for minimally invasive surgery is clearly below that for open surgery, with an almost fourfold increase in the risk of recurrence or death (HR 3.74).

Other relevant outcomes in surgical trials include functional measures, HRQoL, patient experiences and health service and societal parameters, such as return to work and healthcare costs. Surgery-specific outcomes include:

- measures of the extent of resection (for example, positive margin rate, Table 5.1)
- surgical morbidities, such as infection, prolonged bleeding, impaired nerve function and damage to anatomic structures or organs
- 30-day mortality.

Quality assurance in surgical trials can be challenging. Surgical training has changed from an apprenticeship model (observing a senior surgeon then taking on incremental responsibility for components of the procedure) to using simulation and competency-based training and robotics. When a new procedure is introduced, there is both a technological and personal learning curve for the individual surgeon.

To implement quality assurance in surgical trials, participating surgeons can be required to perform a minimum number of procedures using the trial interventions before they are able to recruit subsequent patients to the study. Alternatively, assessment is made of the first few trial patients and then a decision is made for the surgeon to continue. Sometimes the operations (or a sample of them) are video recorded, and then reviewed centrally to check that they were performed acceptably; however, this can be expensive and cumbersome to organize.

Radiation therapy

About half of cancer patients receive RT at some point, and it can be used for any stage of disease. Conformal RT based on high-energy X-rays (photons) has been the standard method for many years. Newer approaches have been designed to deliver a higher dose of radiation more directly and selectively to tumor tissue. These include IMRT, image-guided RT, stereotactic ablative body RT, stereotactic radiosurgery and proton beam therapy.

Total body RT can be used to prepare patients with hematologic malignancies for stem cell transplantation. Internal RT includes brachytherapy (radioactive implants in and around the tumor) and use of radiopharmaceuticals, in which patients swallow or are injected with a radioactive substance.

Novel systemic agents (targeted drugs and immunotherapies) can be combined with sophisticated RT to minimize toxicity to healthy tissue and organs, aiming to improve local tumor control, PFS and OS compared with systemic therapies alone.

Access to RT is a general issue, particularly in developing countries, with shortages of appropriate equipment, radiation oncologists and radiotherapists. A review of all cancer treatment trials registered on clinicaltrials.gov (2007–17) showed that RT trials were less likely than trials of other treatments (mainly drugs) to have national or international sites.[7]

RT trials have several goals, including dose or treatment de-escalation or dose-escalation. Two review articles summarize how new approaches are developed and evaluated in clinical trials.[8,9]

Recruitment. Proton beam therapy has been publicized in the media, so accrual to those trials is unlikely to be difficult. However, advanced technologies may only be available in specialist cancer centers, which limits patient access. Accrual into trials of other types of RT depends on what the trial is evaluating and how the benefits and harms are described to patients. Studies comparing RT with a very different modality (for example, surgery or oral drugs) can be particularly difficult to recruit to because of the clear difference in treatment delivery.

A non-inferiority trial comparing RT with laser surgery (endoscopic incision) for early glottic cancer illustrates how recruitment can fail.[10] The non-inferiority objective was interpreted by several patients as the two interventions having a similar effect on the risk of recurrence, so

they focused on other attributes. Surgery (a single invasive procedure, after which some patients could be discharged home the following day) was clearly preferable to RT, which involved 16–20 visits (fractions) over 3–4 weeks. Travel and the expense of hospital visits were major obstacles. The study was stopped early because so few patients agreed to be randomized.

For early-stage cancers, patients will already have received surgery and probably neoadjuvant and adjuvant systemic therapies, so they often weigh the potential value of RT (as an additional treatment) against any additional clinic visits. These visits may also require patients and their caregivers to travel significant distances to the nearest center and hotel accommodation if RT is to be delivered on consecutive days.

Interventions. Trials may compare different types of RT, with and without other anticancer treatments, or evaluate the timing, duration and dose of RT. Likely comparisons are presented in Table 5.2. Table 5.3 illustrates the breadth of research possible with RT in

TABLE 5.2

Examples of RT interventions that are evaluated in clinical trials

Superiority trial	Non-inferiority trial	Either trial objective
• RT vs no RT (after standard anticancer treatment) • Concurrent RT plus standard systemic therapy vs systemic therapy alone • RT during (or before) vs after systemic therapy • Higher RT dose vs standard dose	• RT vs surgery or chemotherapy • Lower RT dose vs standard dose	• 'New' RT (for example, IMRT or proton beam) vs standard conformal RT • Short course RT (for example, hyperfractionated) vs long course RT

RT can be used to: improve clinical outcomes such as PFS, DFS and OS (for example, dose escalation); maintain these outcomes but with less toxicity (for example, dose de-escalation); or to reduce symptoms and alleviate pain (for example, palliative care).
vs, versus.

TABLE 5.3

Examples of RT trials, illustrating a range of objectives and outcome measures

Patients	Objective	Experimental vs control	Primary endpoint	Conclusions
Metastatic cancer: spread to the spinal canal causing compression (palliation)[11]	Less RT (symptom control): non-inferiority	1 fraction (1 day) vs 5 fractions (5 days)	Mobility status	1 fraction non-inferior to 5 fractions, so patients (end of life) can avoid unnecessary clinic visits
Prostate cancer: spread to the bones[12]	Add RT to SOC (improve prognosis): superiority	Radium-223 vs placebo	OS	Radium-223 reduced the chance of dying by 30%, and increased median OS by ~3 months
Well-differentiated thyroid cancer (curative)[13]	Reduce RT dose (to reduce toxicity): non-inferiority	Factorial trial: 1.1 vs 3.7 GBq (radioiodine ablation) and thyrotropin alfa vs THW	Ablation success 6–9 months after surgery (undetectable thyroid remnant)	1.1 GBq non-inferior to 3.7 GBq; thyrotropin alfa non-inferior to THW 1.1 GBq and thyrotropin alfa could be given on an outpatient basis (instead of 3–5 days in isolation for standard 3.7 GBq and THW)
Pharyngeal cancer (not metastatic)[14]	Advanced RT (to reduce toxicity): superiority	Parotid-sparing IMRT vs standard conformal RT	Severe dry mouth at 12 months	54 fewer cases of dry mouth at 12 months for every 100 patients treated with IMRT vs standard RT
Advanced small-cell lung cancer[15]	Timing of RT in relation to standard CT (improve survival): superiority	Early RT (given concurrently when starting CT) vs late RT (given at the end of CT)	OS	Early RT reduced the risk of dying by 27% if patients were able to complete their course of CT

CT, chemotherapy; GBq, giga becquerels; THW, thyroid hormone withdrawal; vs, versus.

Phase III trials, including a variety of primary endpoints.[11–15] Occasionally, trials are used to evaluate experimental drugs that can act as radiosensitizers to enhance the anticancer effects of RT.

Adherence to RT can be a problem in clinical trials. External-beam RT usually requires several clinic visits over a relatively short period (for example, daily fractions for a few weeks) in order to prevent cancer cells from growing and repopulating between treatments. Cancer patients are often elderly, with comorbidities and fatigue, and the frequency of travel may deter patients from attending sessions. Also, interruptions in RT can have a negative impact on relapse and survival, thus diluting the treatment effect in trials.

Outcome measures. Standard efficacy measures are used in trials of RT, such as OS, relapse/recurrence and PFS. Endpoints relating to toxicities and symptom control are also used, depending on the research question (see Table 5.2 for examples). Long enough follow-up can be an issue (particularly in early-phase trials that are meant to be conducted relatively quickly), as some radiation-related toxicities only appear after several months. Radiation fibrosis (scarring or hardening of tissue) in the skin, subcutaneous tissue and specific organs near to the site of RT (for example, lungs, heart and bladder) is an example of a late effect.

Quality assurance. The delivery of high-quality RT in a consistent manner can be difficult, particularly across geographic sites and with differing levels of expertise among radiation oncologists.[16] Although the use of computerized planning may improve the accuracy and quality of the treatment, the introduction of advanced methods such as IMRT and treatment contouring have made this increasingly complex.

Contouring further highlights the potential interobserver variability in determining target fields. The use of real-time radiation review may be possible in small trials, but it might be too expensive and time-consuming for a large multicenter trial. However, a well-designed trial may not meet its objectives if the RT regimen in each arm is not sufficiently consistent and accurate.

Most trials using RT now include, as a minimum, a review of the treatment plan and its compliance with the protocol, a review of target volumes by at least one other radiation oncologist, and the use of benchmarking cases. Detailed treatment plans for some or all patients could be reviewed centrally.

Cell and gene therapy

Cell therapy involves taking a sample of a patient's cells, manipulating them in the laboratory (and in some cases modifying them genetically) before transferring them back to the patient with therapeutic intent. For cellular immunotherapy, the patient's own immune cells are modified, often to enable them to recognize and attack cancer cells. Gene therapy aims to modify or manipulate the expression of an individual gene or correct abnormal genes in the patient's cells. Cell and gene therapies are sometimes referred to as 'advanced therapies'. One example, chimeric antigen receptor (CAR) T-cell therapy, is a major gene cellular immunotherapy for relapsed diffuse large B-cell lymphoma and acute lymphoblastic leukemia. While high remission rates are seen in some cancer patients treated with advanced therapies, others do not respond at all or their disease relapses. Identifying these patients with poor prognosis is often a key focus for newer trials of gene and cell therapies.

Clinical trials of advanced therapies tend to be conducted in a few expert centers with access to highly specialized manufacturing laboratories. Some countries have particular regulations and guidelines covering advanced therapy trials (which are currently classified as 'high risk'), requiring close monitoring of patients for safety and mortality during the trial and for up to 10 years after.

Recruitment. Because cell and gene therapies are generally evaluated in patients with poor prognosis, accrual into clinical trials has been generally good, particularly for childhood cancers.

Design. The majority of advanced therapy cancer trials have so far been early-phase single-arm (not randomized) studies involving relatively few patients (20–100).[17,18] More recently, randomized Phase II designs have been used, occasionally with a 2:1 allocation ratio, which can make the study more appealing to patients who see this as

a two-thirds chance of getting the experimental therapy (instead of a 50% chance with 1:1 allocation). The same design principles described in Chapters 2 and 3 can apply.

Outcome measures. Efficacy outcomes are similar to those used in trials of other interventions, although advanced therapy trials are not usually powered to evaluate relapse and death rates, which might take too long in some circumstances. Other important endpoints include:

- persistence and duration of circulating CAR T cells (or other cell products) in the blood
- tumor response (the aim is complete remission)
- the ability to collect cells and manufacture enough product for the patient
- neurotoxicity and cytokine release syndrome (a form of systemic inflammatory response)
- a variety of biomarkers for response and toxicity, usually measured in blood.

Conducting advanced therapy trials. Specific issues and challenges include the following.

- The manufacture of cell therapy products is expensive and labor intensive, although more automated systems are being developed, and third-party 'universal' approaches are being investigated.
- Sophisticated laboratory manufacturing processes need to be in place before the trial starts (receipt of cells, storage, production and transport back to patients); they must meet Good Manufacturing Practice (GMP) requirements (see Chapter 6).
- Many of the cancer types are uncommon, so production slots in the manufacturing laboratory could be limited by the time an eligible patient has been identified and consented to the study (so patients are, instead, treated with conventional therapies to avoid further delaying cancer treatment).
- Multiple serial blood sampling is often required for efficacy and safety endpoints (sometimes bone marrow when feasible).
- Non-standard laboratory tests are required for some outcome measures.
- Because only a few cancer centers can conduct these trials, patients and their carers may have to travel long distances if they wish to participate, although some follow-up visits after treatment has finished can be done at a local hospital.

Key points – trials of non-drug interventions

- Clinical trials assessing surgery and technologies such as RT have unique challenges compared with drug trials, including trial design, quality assurance and minimization of bias.
- The way in which surgical and RT trials are described to potential patients (and surgeons) requires care, because patients will consider several features in addition to potential efficacy benefits, such as treatment duration and short- and long-term side effects.
- Quality assurance for these trials should be systematic and specified in advance.
- Trials of advanced cell or gene therapies tend to be small, and they require careful oversight and monitoring of patients; they are usually conducted in expert cancer centers.

References

1. Chapman SJ, Shelton B, Mahmood H et al. Discontinuation and nonpublication of surgical randomised controlled trials: observational study. *BMJ* 2014; 349:g6870.

2. McCulloch P, Cook JA, Altman DG et al. IDEAL framework for surgical innovation 1: the idea and development stages. *BMJ* 2013;346:f3012.

3. Cook JA, McCulloch P, Blazeby JM et al. IDEAL framework for surgical innovation 3: randomised controlled trials in the assessment stage and evaluations in the long term study stage. *BMJ* 2013; 346:f2820.

4. Ramirez PT, Frumovitz M, Pareja R et al. Minimally invasive versus abdominal radical hysterectomy for cervical cancer. *N Engl J Med* 2018;379:1895–904.

5. Treasure T, Baum M. An approach to randomization into surgical clinical trials. *Br J Surg* 2017;104: 11–12.

6. Chagpar AB, Killelea BK, Tsangaris TN et al. A randomized, controlled trial of cavity shave margins in breast cancer. *N Engl J Med* 2015;373:503–10.

7. Liu X, Zhang Y, Tang LL et al. Characteristics of radiotherapy trials compared with other oncological clinical trials in the past 10 years. *JAMA Oncol* 2018;4:1073–9.

8. Thompson MK, Poortmans P, Chalmers AJ et al. Practice-changing radiation therapy trials for the treatment of cancer: where are we 150 years after the birth of Marie Curie? *Br J Cancer* 2018;119:389–407.

9. Sharma RA, Plummer R, Stock JK et al. Clinical development of new drug-radiotherapy combinations. *Nat Rev Clin Oncol* 2016;13:627–42.

10. Hamilton DW, de Salis I, Donovan JL, Birchall M. The recruitment of patients to trials in head and neck cancer: a qualitative study of the EaStER trial of treatments for early laryngeal cancer. *Eur Arch Otorhinolaryngol* 2013;270:2333–7.

11. Hoskin PJ, Hopkins K, Misra V. Effect of single-fraction versus multifraction radiotherapy on ambulatory status among patients with spinal canal compression from metastatic cancer: the SCORAD randomized clinical trial. *JAMA* 2019;322:2084–94.

12. Parker C, Nilsson S, Heinrich D et al. Alpha emitter radium-223 and survival in metastatic prostate cancer. *N Engl J Med* 2013;369:213–23.

13. Mallick U, Harmer C, Yap B et al Ablation with low-dose radioiodine and thyrotropin alfa in thyroid cancer. *N Engl J Med* 2012;366: 1674–85.

14. Nutting CM, Morden JP, Harrington KJ et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a Phase 3 multicentre randomised controlled trial. *Lancet Oncol* 2011;12:127–36.

15. Spiro SG, James LJ, Rudd RM et al. Early versus late radiotherapy as combined modality treatment for limited-disease small-cell lung cancer: a London Lung Cancer Group multi-centre randomised clinical trial. *J Clin Oncol* 2006;24:3823–30.

16. McDowell L, Corry J. Radiation therapy quality assurance in head and neck radiotherapy – moving forward. *Oral Oncol* 2019;88:180–5.

17. Abou-El-Enein M, Hey SP. Cell and gene therapy trials: are we facing an 'evidence crisis'? *EClinicalMedicine* 2019;7:13–14.

18. Ginn SL, Amaya AK, Alexander IE et al. Gene therapy clinical trials worldwide to 2017: an update. *J Gene Med* 2018;20:e3015.

6 Setting up and conducting trials

Clinical trials require careful planning and involve many people with the complementary expertise required to deal with the numerous processes, regulations, governance, and guidelines and recommendations for 'best practice'. Many pharmaceutical companies employ a commercial contract research organization (CRO) to set up and conduct the trial on their behalf. Non-commercial organizations (for example, universities and academic cancer centers) usually conduct the trial themselves. Investigator-initiated trials involve collaborations between academic institutions (the sponsor) and pharmaceutical companies (who provide the drug or drugs and sometimes funding), when the trial concept originated from the academic investigators.

Figure 6.1 is an overview of the key documents, procedures and approvals required for trial set up and conduct, covered in the sections below. Investigators and sponsors also need to ensure that they have sufficient funding in place for the planned trial. For non-commercial studies this usually involves applying for grants from a range of national and international governmental, public sector and charitable organizations.

Critical roles

The sponsor is any organization that has ultimate responsibility for the design, conduct and financing of the trial, and ensuring that the study is conducted in accordance with relevant regulations and guidelines. A sponsor can be the manufacturer of a licensed or unlicensed drug (pharmaceutical company), or a university, hospital or healthcare facility. For international studies, the sponsor can take on all the responsibilities for the trial in all countries or delegate them to a lead institution in each country.

The chief, principal or lead investigator is a health professional named as the lead investigator on applications for regulatory and

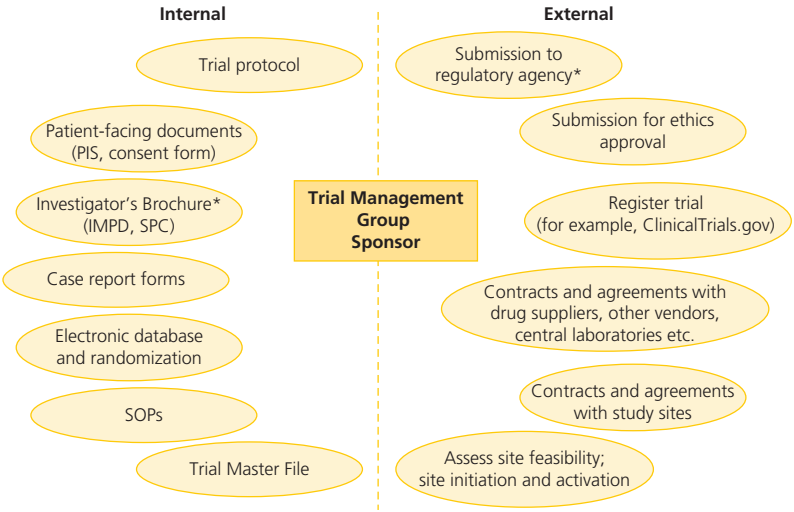


Figure 6.1 Overview of the requirements for setting up a clinical trial, internal and external to the sponsor. *Required for trials of investigational drugs and some medical devices. IMPD, Investigational Medicinal Product Dossier; PIS, Patient Information Sheet; SOPs, standard operating procedures; SPC, Summary of Product Characteristics.

ethical approval, with oversight of the whole trial. They may have conceived the trial proposal, be a key opinion leader in the disease area or a senior employee of the sponsor. This person has to review serious AEs and can be the first point of contact for major problems that arise.

The principal site investigator is a single named health professional who is responsible for conducting the trial at a particular center (site).

Essential requirements for investigator roles. The chief and principal site investigators must be appropriately qualified and currently registered as health practitioners, with certified knowledge and training in Good Clinical Practice (GCP). They must be designated before trial sites can be activated.

The Trial Management Group or **Trial Steering Committee** typically consists of 6–10 people with different areas of expertise necessary to design and conduct the trial: clinicians, scientists, statistician, pathologists, study coordinator, health economist and patient

representatives. This group, usually chaired by the chief investigator, has regular oversight of the study, reports to the sponsor and is responsible for changes to the design and for identifying and addressing problems such as poor accrual. The group also develops a detailed budget for the trial, and in many cases submits a grant application for funding.[1]

Regulations and guidelines

The International Conference on Harmonization (ICH) guidelines provide core principles for trial conduct.[2] These include GCP (ICH-GCP) to ensure the safety and wellbeing of trial patients and the integrity of trial data and GMP (ICH-GMP) to ensure the high quality of drugs (Table 6.1).

Regulations vary between countries and can cover the following aspects of clinical trials:

- trial participation, consent and safety
- independent ethics review of the planned research
- data protection and confidentiality
- storage and use of human tissue
- special patient groups (for example, limited mental capacity and vulnerable adults, children and adolescents).

TABLE 6.1

Main features of GCP and GMP

Good Clinical Practice	Good Manufacturing Practice
• Informed consent from patients must be obtained • Potential risks to patients should be outweighed by the possible benefits • Patients' rights, safety and wellbeing prevail over the interests of science and society • The design should be scientifically sound, with quality and accuracy of the data • The trial should be conducted by appropriately qualified health professionals	• Manufacturing facilities and equipment are constructed appropriately • High-quality production processes are in place, with clear documentation, validations and regular audit and quality control • Staff are appropriately trained and kept safe

These regulations govern trials of drugs, cell and gene therapy and some medical devices. The main European Union regulation is the EU Clinical Trials Directive (2001/20/EC), which includes its own GCP (EU-GCP) guidelines, and is applicable to all EU member states.[3] EU-GCP significantly overlaps with ICH-GCP but has fewer requirements in some parts. However, ICH-GCP may be preferred by commercially sponsored studies because it is more widely accepted by regulatory agencies.

In the USA, the regulations come from the FDA (for example, the Food, Drug, and Cosmetic Act and the Code of Federal Regulations).[4] Access to personal data is also regulated in countries by, for example, the General Data Protection Regulation (GDPR) in the EU.

Trials that include biomarkers in blood, tissue or urine to define primary biological endpoints, or as exploratory translational research, may follow Good Clinical Laboratory Practice (GCLP) guidelines, to ensure the validity and reliability of the laboratory techniques, assays and reagents. GCLP also outlines how biological specimens are processed and stored.

Fewer regulations apply to trials of interventions such as surgery and RT; but essential regulations include the requirement to inform patients about a trial and obtain their consent to participate, data protection and approval by an independent ethics committee.

Documentation

The trial protocol is an essential single document that aims to ensure the trial is conducted consistently and to the same standard at all sites. It is used to demonstrate that regulatory guidelines and recommendations for good practice are met. Approval of the trial protocol by the national regulatory agency (for trials of drugs and some medical devices) and local ethics committees (all trials) is required before any patients can be recruited. The main sections in a protocol correspond to the features shown in Table 1.2 (see pages 9 and 12).

The Patient Information Sheet (PIS) provides information about the trial so that patients are fully informed before giving consent to participate. The PIS describes, in patient-appropriate language:

- the experimental and control interventions
- what patients will be required to do (for example, clinic visits/investigations)

- the possible benefits and a list of the expected adverse effects of the interventions
- statements relating to data confidentiality and protection, and who will have access to and use of their data and biological specimens
- the procedures that will be followed if a patient is harmed (liability and compensation), and who to contact if something serious happens or the participant has any concerns.

Consent form. Having read the PIS, patients who wish to participate in the study must sign and date a consent form (on paper or electronically via a notepad/laptop), which states that the patient has understood the PIS and is willing to proceed. Consent may also be provided by a legal representative of the patient (parent, close relative or guardian) for trials of childhood cancers. The consent form is co-signed by an authorized staff member. Informed consent is a legal requirement.

Case report forms (CRFs) ensure that data on efficacy, adherence and safety are collected systematically and consistently. Some trials use paper CRFs that are completed for each patient by research staff and posted to the central coordinating center. The data are then manually entered into a secure electronic database. Alternatively, electronic data capture systems (eCRFs) allow trial personnel at study sites to enter data directly into the study database. CRF data may come from hospital records, assessments and investigations performed during the trial, while PRO questionnaires are self-completed by patients or through face-to-face interviews.

External data sources. In some cases, outcomes (for example, deaths and recurrences) and other information can be obtained from regional or national registries, insurance/claims databases and prescription databases. The use of personal electronic devices and mobile telephone applications is increasing, allowing real-time collection of data, particularly PROs, directly from patients.

Standard operating procedures (SOPs) ensure that trials are conducted to a consistent standard within the coordinating center and across study sites. SOPs outline the key processes required to set up

and manage the trial. SOPs are used to train new staff, and they also show an external auditor or regulatory inspector that clear and robust systems are in place. Examples of SOPs include how to:

- write the protocol (using a generic template)
- assess the feasibility of sites and activate them
- randomize patients
- record and report AEs, protocol violations and serious breaches
- close down a trial.

Trial Master File (TMF). The collection, curation and stewardship of high-quality data requires complete documentation in the form of a TMF. This may be paper versions of all the key documents kept in a single physical location or it may be stored electronically (an eTMF). The TMF is continuously updated during the trial. At the central coordinating center, it includes the Investigator's Brochure (IB) or equivalent (for drug trials), all protocol versions, indemnification and insurance documents, regulatory and ethics submission and approval documents, signed contracts and agreements and curriculum vitae for all investigators. At study sites, the TMF includes a staff delegation log (listing everyone involved in trial conduct), all communications with the sponsor, and serious AE reports.

The Investigator's Brochure provides detailed information about each drug being evaluated in the trial, including its chemical, toxicological and pharmacological properties, and dose and method of delivery, with supporting evidence from laboratory studies. The IB is developed and updated by the drug manufacturer. For drugs that already have a marketing authorization, European regulatory agencies may accept a Summary of Product Characteristics (SPC or SmPC) instead of a detailed IB, provided that the drug is to be used within the terms of its license.

In Europe, information is required on the quality and safety of all drugs to be used in the trial, including placebo and comparator therapies, within a single document (Investigational Medicinal Product Dossier [IMPD]). This can cross-reference to the IB where appropriate to avoid duplication. If a drug trial has a non-commercial sponsor (for example, a university), the IB, SPC or IMPD is provided by the pharmaceutical company manufacturer.

Agreements and legally binding contracts are expected between the sponsor and organizations involved in the trial. Clinical trial site agreements with each recruiting site/hospital list the roles and responsibilities of the sponsor and local investigators, and include statements relating to insurance and indemnity if a patient is harmed by participating in the trial (for example, due to negligence by the hospital or a problem with the protocol). Other agreements cover drug manufacturing and distribution (drug supply agreement), laboratory testing for translational research including the receipt, handling and storage of biological specimens (material transfer agreements), and any other third-party service referred to in the trial protocol.

Ethics approval

All trials must be approved beforehand by an independent formally recognized organization – an Institutional Review Board (IRB) or Research Ethics Committee (REC). These organizations could be within a hospital or university, or a regional or national agency. Approval is required in each country where patients will be recruited. The IRB or REC is a panel of scientific experts and lay members who examine the trial protocol and all patient-facing documents to determine that there is a good rationale for the study, the design is appropriate, the information is clear for lay people and there are no obvious ethical issues.

Regulatory approval

For trials evaluating a drug (regardless of whether it is already licensed), sponsors must obtain approval from the regulatory agency in each country where patients are to be recruited (see Table 7.1), via an Investigational New Drug (IND) or Clinical Trial Authorisation (CTA) application (see Figure 1.1). The regulatory agency reviews the preclinical studies, dose-finding studies and toxicology data in humans, the trial protocol and PIS. The trial will be approved if it appears to be sufficiently safe for patients, and the benefit–harm balance is likely to be favorable.

In the EU, a single application for regulatory approval can be submitted to the EMA to cover all member states (Voluntary Harmonization Procedure). This may be preferred for trials planned

across several member states, though the central review can be longer than for individual regulatory agencies.

Timelines for ethics and regulatory approval each have a fixed number of days within which a decision must be given or more information requested (typically 30–60 days from submission). Trials conducted in the EU must be registered on the EU Drug Regulating Authorities Clinical Trials (EudraCT) database to obtain a unique EudraCT trial number, before regulatory and ethics approval can be sought.

Changes to the protocol and substantial amendments. Ethics or regulatory organizations can request changes to the protocol and patient-facing materials before giving approval, and they also have to review and approve any major changes proposed by investigators during recruitment or follow-up before the changes can be implemented.

Manufacturing and distribution of trial drugs

The sponsor (usually the manufacturer) will have documentation detailing drug production and compliance with GMP (see Table 6.1) and procedures for labeling and shipping. When hospitals use their own stocks to supply licensed drugs as trial treatments, their pharmacies must have systems to record when trial drugs are dispensed to patients, including identifiers such as batch numbers.

In the EU, manufacturers or importers must hold a manufacturing authorization (from an EU regulatory agency) allowing them to produce, import and label medicines. A single named individual – the Qualified Person (QP; usually an employee of the manufacturer or sponsor) – is responsible for the quality of the drug, in accordance with GMP. The QP signs a release certificate before each batch of drug can be shipped to trial sites. QP release is not required for trial drugs that are already licensed and available from the hospital pharmacy, but it is usually required if the sponsor intends to make changes to the packaging or labeling (for example, blinding).

Registering the trial

All trials should be recorded on an internationally recognized register; the two main ones are ClinicalTrials.gov and the ISRCTN registry. The online listing contains an overview of the objectives and design, expected recruitment timelines and sponsor name.

Registration ensures transparency of all interventional trials and allows future checks on whether the study has been published – failure to publish may raise concerns. Investigators can also see whether trials similar to their own have been conducted or are ongoing. Trials must be registered before the first patient is recruited, otherwise journals may not even consider a submitted trial article.

Conducting the trial

Figure 6.2 shows the main trial activities. Large trials may be started at a few centers first to assess the feasibility of recruitment and identify and correct any issues that arise with trial conduct.

Safety monitoring. Monitoring and reporting AEs are essential functions (called pharmacovigilance in trials of drugs or medical devices). An AE is any disorder or symptom that affects the patient's health or wellbeing, including abnormal biochemical tests and the ability to function. AEs may or may not be causally related to the trial interventions. Events that are judged to be caused by the study treatment are referred to as adverse reactions or adverse drug reactions. AEs should not include cancer progression or death from the cancer of interest.

AEs can be expected or unexpected. Expected events for a particular therapy are specified in the prescribing information of a licensed medicine, or in the IB if not licensed, and in the trial protocol.

Serious AEs (or reactions) (SAEs or SARs) are defined if any of the following criteria are met:

- death
- life-threatening
- hospitalization, or prolonged stay if already in hospital
- results in persistent or significant disability or incapacity
- results in a congenital abnormality or birth defect
- leads to any other condition judged significant by the clinician.

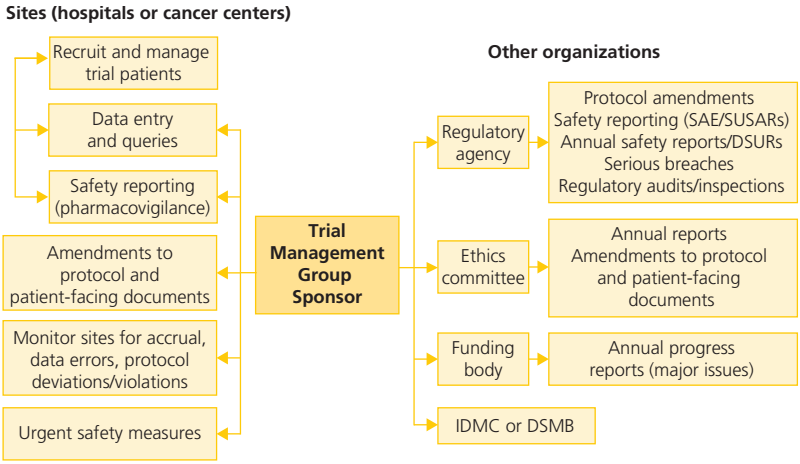


Figure 6.2 Overview of trial conduct. Contact with the regulatory agency is only required for drug trials. DSMB, Data and Safety Monitoring Board; DSUR, development safety update report (evaluation of safety information collected during a specific reporting period, only for a particular trial drug); IDMC, Independent Data Monitoring Committee.

SAEs should be reported to the sponsor (or coordinating center) within 24 hours of discovery. Suspected *unexpected* serious adverse reactions (SUSARs) are judged to be causally related to the study drug and require special processing. Fatal or life-threatening SUSARs must be reported to the regulatory agency by the sponsor, typically within 7 days of being notified by the site. If the SUSAR is not fatal or life-threatening, the regulatory agency is informed within 15 days. The REC or IRB that approved the trial protocol is also notified.

An urgent safety measure (USM) can be implemented immediately if unacceptable AEs are seen during the trial, or evidence obtained elsewhere suggests potential major harm to patients using the current protocol. The revised protocol does not need to go through the standard regulatory and ethics review procedures, but these bodies are informed of the changes. Examples of USMs are lowering the drug dose or adding a major eligibility criterion.

Monitoring study sites

Monitoring is an oversight function performed by the sponsor or its delegate, the level of which depends on the complexity of a trial and

potential risks to patients. Monitoring can be done centrally (by the coordinating center) using the trial database or by the trial team who visit sites and access patient files and other documentation (on-site monitoring). This can include:

- checking that signed consent forms were obtained from participants
- checking trial data in the patients' records against the CRF and trial database, to identify data entry errors (source data verification)
- looking out for major protocol violations and deviations, which are instances where the protocol has not been followed (for example, ineligible patients recruited in error, important assessments are missing, or dosing errors of the trial drug or RT); these may or may not have an impact on patients or the validity of the trial data
- checking that SAEs and SUSARs have been reported and within the required timelines
- looking for patterns in AEs across the trial that might cause concern.

Audit and inspection

Many countries have a system for inspecting the offices of the sponsor, the trial coordinating center, one or more study sites, the drug manufacturing facilities or a laboratory responsible for biomarker testing, particularly if the marker is integral to the design or primary endpoints of a trial. These inspections are performed for drug trials by the national regulatory agency and can be preplanned or triggered by an unexpected and urgent serious concern. A single trial or several trials can be inspected during a visit that often lasts 3–5 days. The inspectors check that:

- all necessary regulatory and ethics approvals and signed agreements have been obtained
- all trial documentation (for example, the TMF) is available, complete and up to date and insurance policies are in place
- GCP and GMP guidelines are followed
- there are clear systems for monitoring safety, and SAEs and SUSARs are reported on time.

Inspectors can suspend or stop a trial if they find serious issues, especially any that compromise patient safety. For trial interventions other than drugs, the sponsor may decide what level of independent audit is useful, and this function can be performed by staff internal or external to the sponsor's organization.

Ongoing independent review and monitoring

An Independent Data Monitoring Committee (IDMC) or Data and Safety Monitoring Board (DSMB) is a group of 3–5 individuals, including health professionals and a statistician, with no direct connection to the clinical trial. They provide an independent review of recruitment, data completeness, protocol violations/deviations and any general problems with the study. During a 'closed' (confidential) session, the IDMC members also examine detailed data on AEs, adherence and efficacy according to trial arm (sometimes even for double-blind trials). The primary purpose of the review is to ensure that the AE profile is acceptable and, when efficacy data are also available, that there is a favorable risk–benefit balance. The IDMC can recommend changes to the trial protocol or conduct, or stop the study early because of unacceptable AEs due to the experimental treatment; other reasons for stopping are shown in Box 6.1.

Statistical analysis plan

The statistical analysis plan (SAP) details how trial data will be handled and presented, and the specific statistical analyses that will be performed at the end of the trial, including prespecified subgroup analyses. The document is finalized before the database is ready for full analysis; one reason for this is to avoid multiple unplanned analyses that produce spurious treatment benefits ('data mining'). It also describes how the analysis will handle:

- multiple efficacy endpoints, especially if there are two or more primary endpoints
- patients who withdrew from the trial early or were lost to follow-up
- missing data (where patients have no efficacy measures)
- repeated measures (where an endpoint has been measured several times for each patient).

The SAP would be included with submissions to regulatory and HTA agencies for licensing and market access.

Closing the trial

The formal end of a trial (i.e. the point at which data collection will stop) may not be obvious. It could be after a defined period (for example, 3 years after the last patient has been recruited), or after

BOX 6.1

Reasons why trials do not meet their intended objectives ('failed' or 'negative' trial)

- The objectives are no longer of interest, making the trial out of date
- The control (comparator) treatment is no longer current standard clinical practice
- Poor accrual: lack of interest from sites, or patients unwilling to participate because they do not like the protocol (for example, at least one intervention is unappealing or too many extra clinic visits and investigations are required)
- The event rate (for example, number of deaths for OS) is too low, making the analysis of the primary endpoint unreliable and inconclusive
- For a superiority trial, none of the primary or key secondary efficacy endpoints show a statistically significant benefit with the experimental intervention, including in relevant subgroup analyses
- For a non-inferiority trial, the results for the primary and perhaps other key endpoints clearly overlap with the non-inferiority margin, so it cannot be concluded reliably that the experimental intervention is non-inferior to the control
- Funding has ended so that the sponsor cannot finish recruitment and/or follow-up

a certain number of events has been seen. The definition has implications for funding and study sites. The trial protocol may allow key long-term data collection during an observational phase, particularly if data are collected electronically from patient records or regional or national registries.

Why trials 'fail'

The terms 'failed trial' or 'negative trial' are unsatisfactory but frequently used. They imply that the trial generated no new information and perhaps even wasted resources, although this is rarely the case. Box 6.1 gives some likely reasons for 'failure', though

a better phrase is possibly 'did not meet its objectives'. A trial that had too few deaths for a primary endpoint of OS to be statistically significant, but generated enough events for PFS (or DFS) assessment, may be difficult to interpret, but in some cases decision-makers may still change practice based on PFS/DFS, particularly if there is supporting evidence from elsewhere. The perception of a 'failed' trial is one of the main reasons for publication bias (researchers are less likely to submit, and some journals are less likely to review or accept for publication); all such trials should be published.

Key points – setting up and conducting trials

- National regulations apply to all types of trials, to ensure patients are adequately informed about the trial before they participate, and to assure their safety and data confidentiality.
- Ethics approvals are mandatory in all countries where patients will be recruited, and regulatory approval is required in addition for trials of drugs and some medical devices.
- Many documents are needed when conducting a trial, and are contained in the TMF. The protocol is a major document setting out how the trial will be conducted.
- Monitoring and reporting AEs, particularly SAEs and SUSARs, is a key function.
- Trials that do not meet the planned primary objectives still contribute useful data; the results should be published.

References

1. Hackshaw AK. *How to Write a Grant Application: For Health Professionals and Life Sciences Researchers*. Oxford: Wiley, 2011.

2. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *Efficacy Guidelines*. www.ich.org/page/efficacy-guidelines, last accessed 2 June 2020.

3. European Commission. *Clinical trials – Directive 2001/20/EC*. https://ec.europa.eu/health/human-use/clinical-trials/directive_en, last accessed 2 June 2020.

4. US Food and Drug Administration. *Regulations: Good Clinical Practice and Clinical Trials*. www.fda.gov/science-research/clinical-trials-and-human-subject-protection/regulations-good-clinical-practice-and-clinical-trials, last accessed 2 June 2020.

7 Publishing trial results, changing clinical practice, and supporting evidence

The findings from clinical trials are used in various ways (Figure 7.1). This chapter provides an overview of publishing journal articles, licensing of new treatments and market access, and the use of real-world evidence and meta-analysis as supporting evidence.

Reporting and publishing trial results

For drug trials, there is a legal requirement for at least a summary of the results to be published, which can be on the website of a national regulatory agency (for example, EudraCT for EU trials) or an international trials register such as ClinicalTrials.gov. This should be within 1 year of trial closure for EU studies or the final data collection for the primary outcome for US studies, but there may be valid reasons why a longer time is needed. Beyond legal requirements, there are ethical, moral and scientific obligations for the results and conclusions of all interventional trials to be made publicly available.

There is a skill to writing a clinical trial publication that has a succinct description of the main design features, clear and defendable scientific messages, and conclusions consistent with the results.[1] Journal editors and peer reviewers examine the trial design, analyses and trial conduct (often to look for major problems that affect the reliability of the conclusions), but they also consider whether the trial

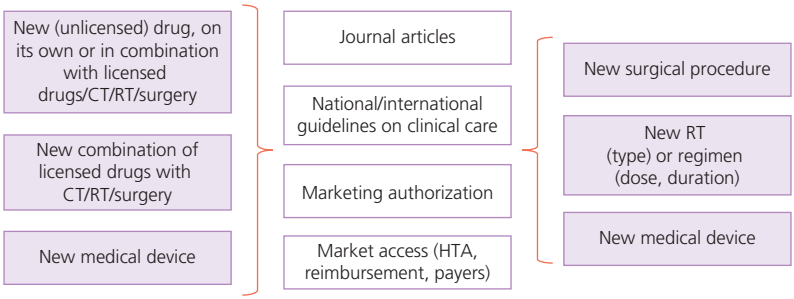


Figure 7.1 Use of clinical trial results. CT, chemotherapy.

has added new information to what is already known. Authors sometimes exaggerate the benefits of a new intervention or underplay the harms. It is therefore important that health professionals understand trial design and can interpret quantitative findings themselves, without relying on the authors' opinion in the abstract and paper.[2]

Many journals expect articles to be written in line with the Consolidated Standards of Reporting Trials (CONSORT) guidelines.[3]

As well as the distinction between primary and secondary endpoints, there are also the concepts of exploratory endpoints that are prespecified (included in the protocol and SAP) or post hoc (unplanned), which require justification.

Phase III trials that meet their primary objectives may influence clinical guidelines such as those developed by the US National Comprehensive Cancer Network (NCCN), American Society of Clinical Oncology (ASCO) and European Society for Medical Oncology (ESMO).

Marketing authorization

Clinical trials are the most important source of evidence to determine whether a newly developed drug (or medical device) can be introduced into clinical practice. The manufacturer requires a marketing authorization (i.e. a license) to market and distribute a new pharmaceutical product for use in the geographical jurisdiction of the awarding agency.[4,5] Most countries have a regulatory agency (Table 7.1) to conduct a comprehensive review of the efficacy and safety data generated from clinical trials alongside other supporting data (for example, pharmacological information). This is to ensure that the potential benefits to patients outweigh any harms (Figure 7.2). These agencies also set standards and guidelines for the conduct and reporting of clinical trials.

The license details how the drug is to be used (for example, the line of treatment, as monotherapy or in combination with other specified drugs, and the dose and route of administration) and in which specific patients (for example, tumor type and stage of disease, and possibly defined further by a biomarker). The regulatory agency must also authorize any changes to the formulation (for example, ingredients, mode of delivery).

TABLE 7.1

Examples of national regulatory and HTA agencies

Region	Regulatory agency	HTA
USA	Food and Drug Administration (FDA)	Several, for example, Agency for Healthcare Research and Quality (AHRQ); Medicare, Medicaid and private insurance providers
Canada	Health Canada	Canadian Agency for Drugs and Technologies in Health (CADTH)
Europe	European Medicines Agency (EMA)	
UK	Medicines and Healthcare products Regulatory Agency (MHRA)	National Institute for Health and Care Excellence (NICE)
Germany	Federal Institute for Drugs and Medical Devices (BfArM) and the Paul-Ehrlich-Institut	Institute for Quality and Efficiency in Healthcare (IQWiG)
France	National Agency for the Safety of Medicines and Health Products (ANSM)	Haute Autorité de Santé (HAS)
Japan	Pharmaceuticals and Medical Devices Agency (PMDA)	Ministry of Health, Labour and Welfare (MHLW)
Australia	Therapeutic Goods Administration (TGA)	Pharmaceutical Benefits Advisory Committee (PBAC)

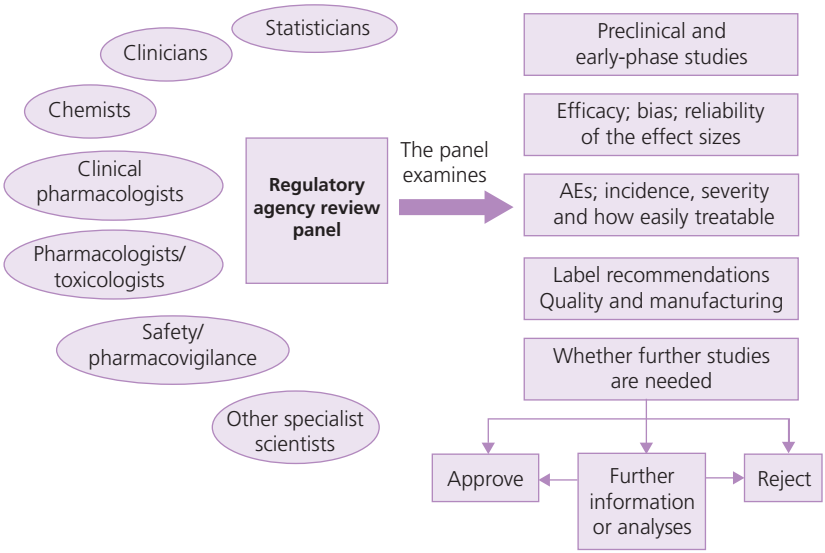


Figure 7.2 Typical regulatory review process to provide a marketing authorization.

In the EU, manufacturers can apply to the EMA for a single marketing authorization that covers all member states, or to national agencies separately. The central procedure may be more efficient in the long term, but applying for licenses in individual countries can allow a more focused strategy with review times that are shorter than with the EMA central procedure.

Standard review processes typically take 10–12 months, but regulatory agencies also offer several alternative review pathways, including accelerated approval (FDA) or conditional authorization (EMA), based on preliminary data (for example, Phase II) and may be contingent on the subsequent submission of mature survival data. Quicker processes (for example, fast track and breakthrough designation) are also available for highly innovative drugs, new drugs providing a major advance in treatment and drugs for rare diseases. Phase II data may be accepted for initial authorization in these scenarios. These alternative review pathways could take 60 days to 8 months from submission to final decision.

Ideally, at least two trials of the same intervention and cancer type (usually in different geographical locations) are required, to ensure that the therapy is effective and the magnitude of the benefit is reliably estimated. However, a single study may be accepted for uncommon tumor types, areas of high unmet need or when efficacy results are remarkable.

For cancer drugs intended for children and adolescents, the sponsor would provide sufficient data on efficacy and safety after following a Paediatric Investigation Plan (PIP) in Europe, or a Pediatric Study Plan (PSP) in the USA, that had been developed as part of Phase I–III trials.

Market access (reimbursement)

Market access refers to the process by which a new treatment (single drug, new combination or new indication) is introduced into routine practice so that it is accessible to all eligible patients at an affordable price.[6] Many countries have an HTA or reimbursement agency (Table 7.1) that evaluates the clinical efficacy and safety data, the extent of unmet need in the intended population, value for money given the proposed drug cost, and any impact on healthcare delivery. HTA agencies or the healthcare provider may set or negotiate the price or decide whether (or to what extent) a drug is reimbursed or funded, which may include recommending the drug only in a defined subgroup of patients.

Market access covers three linked areas:

- HTA: detailed evaluation of efficacy, safety, HRQoL and cost
- pricing and reimbursement rate (P&R) from the healthcare provider
- formulary listing: inclusion of a drug on national, regional or hospital lists of approved medicines.

Several surgical procedures and RT may also undergo an HTA review in some countries, but the evaluation can be simpler than that for drugs.

Within a commercial company, a clinical team, regulatory affairs professionals, health economists, medical science liaison representatives and patient engagement groups together develop a market access strategy, usually starting from the earliest stages of clinical trials. This aims to ensure that the Phase II and III trials, in particular, are based on an appropriate patient population, include a comparator (control) that is relevant in the target countries, and have endpoints likely to be accepted by HTA agencies.

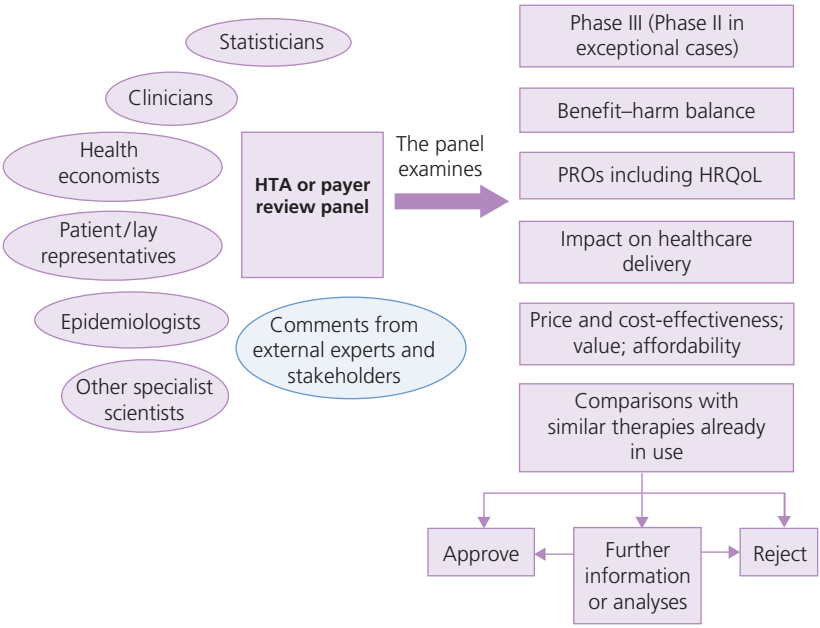


Figure 7.3 Typical HTA review process to provide approval for market access and reimbursement.

While regulatory agencies focus on the efficacy and safety of a new treatment, HTA evaluates its use in the context of the local (regional or national) patient population (Figure 7.3). The market access processes and criteria vary between different countries, including the methods used to evaluate treatments, which can be a challenge for manufacturers, who may have to conduct two or more clinical trials to satisfy several HTA agencies.

The evaluation of trials by HTA agencies differs from evaluation by the regulatory agencies in key areas (compare Figures 7.2 and 7.3). The submission dossiers for HTA agencies need to outline current practice and unmet need in the target population and how this is addressed by the treatment being evaluated.

Patient subgroups. The submission dossier must include data from all patients included in the trial; however, HTA agencies may restrict access to subgroups of patients in whom they consider the greatest

benefit was seen in the pivotal trial, such as only patients with a particular biomarker (for example, PD-L1 expression for an immunotherapy).

Comparator. Regulatory agencies accept placebo as the comparator when ethically acceptable (in addition to some form of SOC, or when no therapies are normally given), which often requires a superiority trial. However, HTA agencies tend to prefer a head-to-head comparison of a new drug versus the SOC in that region if one exists, usually when they are the same class of drug (for example, gefitinib versus afatinib which are both tyrosine kinase inhibitors); this may require a non-inferiority trial.

The choice of comparator used in the clinical trial must be justified in the submission dossier, particularly if it is not the SOC. Sometimes an indirect treatment comparison (see section on network meta-analysis below) is used to compare the new treatment with another therapy if there has been no head-to-head trial.

Endpoints. OS is generally the preferred survival endpoint for advanced cancer (see page 55). HTA agencies can be critical of submissions based solely on PFS data, or if OS data do not show significant benefit because of high crossover rates. Payers may require the patient relevance of any surrogate endpoint to be demonstrated in the tumor type being considered. Payers are particularly interested in the quality of survival as well as its duration, so HRQoL and toxicity data are essential.

Patient convenience. HTA reviewers may also consider other potential positive and negative impacts of a new drug on patients or healthcare delivery if it were approved for routine use. For example, in the treatment of advanced renal cell carcinoma, sunitinib is taken orally once daily at home. This is easier for patients than interferon alfa (the SOC at the time), which is administered as a subcutaneous injection requiring three clinic visits each week.

Health economic evaluation

Clinical trial data are key in the economic evaluation of a new treatment. This type of analysis includes a comparison between two or

more interventions (or with no intervention), a treatment effect on clinically relevant endpoints, and costs. The cost analysis usually focuses on costs to the healthcare provider of introducing the intervention (for example, costs of treatment, additional assessments and investigations, and hospital stay). Economic evaluations may also consider costs to the patient (for example, travel to hospital) and societal costs (for example, number of workdays lost). The analysis will also look at savings (offsets) that could be realized when the new treatment replaces an existing one.

There are several different types of economic analyses, and they can involve complex modeling that requires carefully justified inputs and assumptions. Cost-effectiveness analysis and cost-utility analysis are the most widely used. They produce outputs such as the incremental cost-effectiveness ratio (ICER; the extra cost needed for the additional benefit) and the cost per quality-adjusted life-year (QALY) gained based on clinical efficacy and HRQoL.

Systematic reviews and meta-analyses

Several trials of the same intervention in the same cancer type often yield different findings. A systematic review involves identifying a set of similar trials and combining their results to provide a more reliable and accurate estimate of the treatment effect than any trial on its own. It is considered to be among the highest level of evidence, and organizations such as the Cochrane Collaboration are leading sources of systematic reviews.

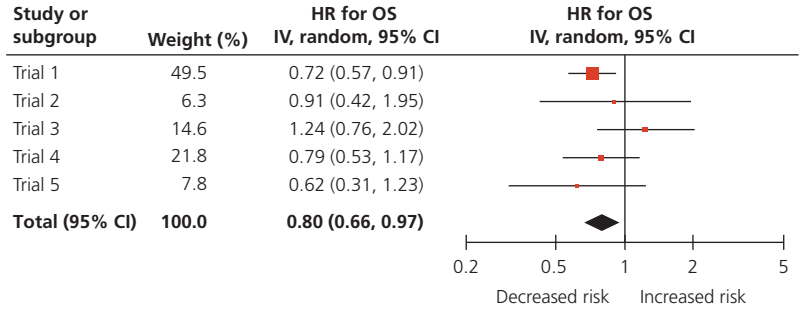
The trials should be randomized and compare the same new intervention with the same comparator across the trials. In reality, the new and control interventions may have different specifications (for example, different doses or treatment durations), and the trial patients are rarely identical. They should be broadly similar, however.

A systematic review involves defining the research question and the criteria for a literature search, and the screening of articles. Each report that meets the inclusion criteria is appraised and relevant summary information extracted (usually effect sizes, such as HRs, sometimes AEs) and an assessment is made of study quality. The statistical process used to combine the results is called a meta-analysis. Several methods are available, mainly fixed or random effects; with the latter, there is an attempt to allow for heterogeneity where the

effect sizes differ noticeably between the trials. Some systematic reviews involve obtaining the raw patient-level data from the trial groups; these are individual patient data (IPD) meta-analyses.

A forest plot is created, which shows the individual effect sizes and the pooled effect size from the meta-analysis (Figure 7.4). Heterogeneity is a major consideration that needs investigation if present.

Systematic reviews and meta-analyses are often used by decision- and policy-makers. They can also be used within submissions to an HTA as supporting evidence for a new intervention for which market access is sought, though the review panel will judge the reliability on a case-by-case basis. Systematic reviews can provide more reliable evidence on subgroup analyses, compared with single trials, because they are based on more patients.



Heterogeneity: $Tau^2 = 0.01$; $Chi^2 = 4.45$; df = 4 ($p = 0.35$); $I^2 = 10\%$
Test for overall effect: $Z = 2.26$ ($p = 0.02$)

Figure 7.4 An example of a forest plot for meta-analyses. Estimates of HR vary and some are statistically significant. A large weight (for example, trial 1) indicates many patients and/or events were seen, increasing the reliability of that study. The size of each square reflects the weight. A single (pooled) measure of treatment effect is estimated using a weighted average of the five effect sizes. The combined effect size here is 0.80 ($p = 0.02$). Two tests for heterogeneity are used to determine if any of the individual trial results are substantially different from the pooled effect: heterogeneity p value 0.35 (we want $p > 0.05$) and $I^2 = 10\%$ (0 indicates no heterogeneity at all, 100 indicates substantial heterogeneity; generally we want $I^2 < 30\%$). In the example, there is little evidence for heterogeneity. df, degrees of freedom; IV, inverse variance method; random, random effects.

Network meta-analysis. This is a specific type of meta-analytic approach that combines efficacy results from several different treatments, even when they have not been directly compared within a randomized trial. An indirect treatment comparison (ITC) is performed. For example, if there are several trials of treatment A versus SOC and others of treatment B versus the same SOC, statistical methods are used to estimate the effect size for treatment A versus B. This can be particularly informative in the absence of a direct head-to-head comparison between two interventions (for example, two drugs in the same class).

A network meta-analysis can combine trials with direct comparisons and ITCs. Manufacturers may include a network meta-analysis in an HTA dossier with the aim of showing that their product is likely to be superior to others or has a similar effect and therefore represents another treatment option, and to support health economic evaluations. However, such analyses are only considered reliable if the trials are sufficiently similar in terms of patient characteristics, delivery of the control treatment and length of follow-up.

Real-world evidence/data

While randomized controlled trials are the most reliable way to assess a new treatment, they have several limitations.

- Patients who participate in trials may not be representative of the target population encountered in clinical practice, because patients who are less fit or have a poor prognosis or comorbidities are often excluded.
- Trials may be conducted at expert cancer centers, with better standards of care and management than in other oncology departments.
- Adherence to the experimental and control treatments tends to be higher in trials than in routine practice because participants are more motivated (and fitter).

The restrictive eligibility criteria and superior adherence may result in the treatment benefit being overestimated, making cost-effectiveness assessments more favorable.

There is no standard definition of real-world data (RWD: the patient-level data collected from various sources) or real-world evidence (RWE: the analysis/synthesis of RWD). The data can be

collected outside of research-intensive or academic environments, such as routine clinical care and home and community settings.[7] Gathering RWD involves the collection of patient and tumor characteristics and patient outcomes (such as death and recurrences) and occasionally PROs, but the number of factors is usually lower than for a clinical research trial. RWE is becoming increasingly useful to address the potential shortcomings of clinical trials outlined above. Table 7.2 lists several uses for RWE.

Sources of RWE. Many real-world studies are large-scale observational epidemiological studies, with long-established design and analysis issues. One source of RWE comes from postmarketing (Phase IV) studies, also known as postmarketing surveillance (pharmacovigilance) studies, conducted by a drug manufacturer after a new drug has been launched in a region or country. This may be voluntary or at the behest of a regulatory or HTA agency. These studies involve the collection of long-term efficacy and safety data, and they may identify rare AEs not seen in the pivotal trials that supported licensing and market access.

TABLE 7.2

Some common uses of RWE

- To examine efficacy and/or AEs associated with a particular treatment in large numbers of people in routine practice
- To provide efficacy data for network meta-analyses to compare two or more therapies in the absence of direct comparisons from randomized trials
- To provide data from patients who had SOC for comparison with data from a single-arm clinical trial of an experimental treatment when large randomized studies are not feasible
- To collect large amounts of data on 'real life' use and experience of new therapies, including adherence and HRQoL (using personal electronic devices and mobile phone applications)

Other sources of RWE are:

- regional and national databases for specific disorders, from general practice, community care or death registries
- electronic health records from clinics, commercially funded databases that use health records (for example, the Flatiron Health Network), insurance claims databases and prescription databases
- specifically designed prospective cohort studies.

Research programs are under way in Europe and the USA to develop acceptable and reliable approaches for RWE (for example, the European Innovative Medicines Initiative and the FDA RCT-DUPLICATE program). One major goal is for RWE to be used as complementary evidence to support licensing[8] and market access, alongside randomized trials.

Organizations that own RWD must ensure data security and patient anonymity, especially when these data are accessed by other organizations or researchers.

Limitations of RWE. Even though RWE is often based on many patients, this does not overcome inherent bias and confounding which are, by design, minimized or avoided in randomized trials.[9] RWE may be unreliable when, for example:

- design and conduct are perceived to be too inferior to a randomized trial
- patients are not sufficiently representative of the target group if the RWE comes from a selected population
- there are methodological/statistical problems, including:
  - missing data for key patient and tumor factors and outcomes
  - major confounding and bias cannot be dealt with adequately
  - there are issues over data quality
- the lack of blinding is likely to have biased subjective endpoints, such as PROs.

The statistical analysis and methods are more complex than for randomized trials, so they must be explained clearly, and different methods can produce different results which may over- or underestimate treatment effects. Sophisticated methods continue to be developed.

Key points – publishing trial results, changing clinical practice, and supporting evidence

- All clinical trials results should be published, regardless of the conclusions.
- New pharmaceutical products require a license (marketing authorization) from a national regulatory agency for routine use, based on a detailed review of the safety and efficacy data, to ensure the benefits outweigh any harms.
- HTA/reimbursement agencies evaluate the value for money of a new treatment based on unmet need in the target patient population, clinical benefits and potential harms, and price (cost-effectiveness in some markets).
- The requirements, processes and focus of regulatory and HTA agencies often differ.
- Systematic reviews and meta-analyses combine information from several similar trials of the same intervention.
- RWE on efficacy and safety is typically collected for large numbers of patients from observational studies and databases of health records from routine practice. This evidence is useful to complement randomized trial data but has inherent limitations.

References

1. Hall G, ed. *How to Write a Paper*, 5th edn. Oxford: Wiley-Blackwell 2012.

2. Greenhalgh T. *How to Read a Paper: The Basics of Evidence-Based Medicine and Healthcare*, 6th edn. Hoboken: Wiley, 2019.

3. CONSORT. *CONSORT Statement*. www.consort-statement.org, last accessed 2 June 2020.

4. US Food and Drug Administration. *Development & Approval Processes: Drugs*. www.fda.gov/drugs/development-approval-process-drugs, last accessed 2 June 2020.

5. European Medicines Agency. *How Are New Medicines Approved by EMA?* www.ema.europa.eu/en/news/how-are-new-medicines-approved-ema, last accessed 2 June 2020.

6. Toumi M. *Introduction to Market Access for Pharmaceuticals*. Boca Raton: CRC Press, 2017.

7. Sherman RE, Anderson SA, Dal Pan GJ. Real world evidence – what is it and what can it tell us? *N Engl J Med* 2016;375:2293–7.

8. US Food and Drug Administration. *Real-world Evidence*. www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence, last accessed 2 June 2020.

9. Collins R, Bowman L, Landray M, Peto R. The magic of randomization versus the myth of real-world evidence. *N Engl J Med* 2020;382:674–8.

Index

Fill the gap in your knowledge, *fast!*

with ***Fast Facts*** – the ultimate medical handbook series

FAST FACTS
Clinical Trials in Oncology

KARGER

ISBN 978-1-912-77673-3

9 781912 776733 >

FAST FACTS

KARGER.COM