HEALTH TECHNOLOGY ASSESSMENT

HOW DO YOU ASSESS CLINICAL TRIALS (IN SURGERY)?

MODULE 4

Workshop Manual
March 2006

Surgery Strategic Clinical Network: Evidence Decision Support Program

This Project was funded by:
Canadian Coordinating Office for Health Technology Assessment (CCOHTA)

WELCOME

Welcome to the **fourth** module of six in a series on Health Technology Assessment (HTA). The primary objective of this fourth module and workshop is to provide you with an overview of clinical trials in general and of how to assess clinical trials in surgery.

We hope that the fundamentals presented in this module will not only assist you in your own assessment of clinical trials in surgery, but also provide you with the tools required to critically evaluate all clinical research in a sound, objective, and appropriate manner.

We look forward to sharing this experience with you and your colleagues. Your feedback and comments on both the module and workshop will be greatly appreciated! Please send comments to the Office of Surgical Research at osr@ucalgary.ca

Prepared by:

Elizabeth Oddone Paolucci

Tyrone Donnon

Paule Poulin

The HTA Education Team Members,

Paule Poulin, PhD
Office of Surgical Research, Associate Director, Calgary Health Region

Tyrone Donnon, PhD
Medical Education & Research Unit, Assistant Professor, Faculty of Medicine

Elizabeth Oddone Paolucci, PhD
Medical Education & Research Unit, Educational Consultant

Norm Schachar, MD, FRCSC
Office of Surgical Education, Director, Department of Surgery

Anita Jenkins, RN, BN, MN
Office of Surgical Education, Education Coordinator, Department of Surgery

David Sigalet, MD, PhD., FRCSC
Office of Surgical Research, Director, Department of Surgery

CONFLICT OF INTEREST

Conflict of interest is considered to be financial interest or non-financial interest, either direct or indirect, that would affect the research contained in a given report, or creation of a situation where a person's judgment could be unduly influenced by a secondary interest such as personal advancement. Indirect interest may involve payment which benefits a department for which a member is responsible, but which is not received by the member personally, e.g. fellowships, grants or contracts from industry, industry sponsorship of a post or a member of staff in the department, industry commissioning of work.

Based on the statement above, no conflict of interest exists with the author(s) and/or external reviewers of the fourth module.

ACKNOWLEDGMENTS

This Health Technology Assessment Module was prepared for the **Department of Surgery** by the **Office of Surgical Research,** the **Medical Education and Research Unit,** and the **Office of Surgical Education** and funded by the **Canadian Office for Health Technology Assessment (CCOHTA)**, a National-based organization with a mandate to facilitate the appropriate and effective utilization of health technologies within health care systems across Canada. CCOHTA's mission is to provide timely, relevant and rigorously derived evidence-based information to decision makers and support for the decision-making processes. The HTA Team wishes to thank Ms. Darlene Bridger and Ms. Linda Davis for their administrative support.

This Project was funded by:
Canadian Coordinating Office for Health Technology Assessment (CCOHTA)

TABLE OF CONTENTS

1.0 OBJECTIVES

The main goal of this Health Technology Assessment (HTA) module is to provide an overview of how to assess clinical trials with a particular emphasis in surgery.

By the end of this HTA module, participants will be able to:

(1) Describe some of the types of clinical trials appearing in the medical literature, specifically:

a. Randomized Controlled Trials

Clinical Trials (Randomized or not)

b. Preclinical Trials
c. Screening Trials
d. Crossover Trials
e. Multi-center and International Clinical Trials
f. Equivalence Trials
g. Screening Trials
h. Safety Trials
i. Explanatory or Efficacy Trials
j. Pragmatic or Effectiveness Trials
k. Blinding or Masking Trials
l. Placebo-Controlled Trials

(2) Identify the purpose and structure of clinical trials.

(3) Discuss the advantages and disadvantages to conducting clinical trials.

(4) Understand the obstacles to conducting clinical trials.

(5) Propose a framework for clinical research in surgery.

2.0 INTRODUCTION

It is a generally accepted principle that the strength of a study depends on its design. Similarly, good design and statistical analysis underpins good-quality work relevant to health technology assessment (White, Ashby, & Brown, 2000). The ability of Health Technology Assessments (HTAs) to answer questions about the effectiveness and cost-effectiveness of new technologies relies on the availability of appropriate methodologies and statistical analyses (White, Ashby, & Brown, 2000). Various hierarchies of evidence have been presented, however, one commonality is that randomized controlled trials tend to be one of the strongest of study designs that produce sound sources of evidence.

Medical practice is changing, and the change that involves using the medical literature more effectively in guiding medical practice is profound enough to be called a paradigm shift (American Medical Association, 1992). The foundations of the paradigm shift lie in developments in clinical research over the last 30 years. In 1960, the randomized clinical trial (RCT) was an oddity. It is now accepted that virtually no drug can enter clinical practice without demonstration of its efficacy in clinical trials (American Medical Association, 1992). Moreover, the same randomized trial method increasingly is being applied to surgical therapies and diagnostic tests (American Medical Association, 1992). Much of the statistical literature on study designs that relate to health technology assessment comes from clinical trials; there are relatively few publications that cover the more complex experimental designs, meta-analysis or studies of drug safety (White, Ashby, & Brown, 2000).

Since Clinical Trials (CTs), in particular randomized clinical trials (RCT) are central to the work of many medical professionals and are believed to provide the most compelling evidence of a causal relationship between treatment and effect, it is important that they be well understood. In the following section we begin our introduction of CTs.

3.0 TYPES OF CLINICAL TRIALS

In medicine, a clinical trial (also known as clinical research) is a research study with human volunteers with the aim of evaluating new drugs, medical devices, biologics, or other interventions to patients in strictly scientifically controlled settings (http://www.clinicaltrials.gov/ct/info/whatis#whatis). Clinical trials are required for regulatory authority approval of new therapies. Trials may be designed to assess the safety and efficacy of an experimental therapy, to assess whether the new intervention is better than standard therapy, or to compare the efficacy of two standard or marketed interventions. The trial objectives and design are usually documented in a clinical trial protocol (http://en.wikipedia.org/wiki/Clinical_trial; http://www.clinicaltrials.gov/ct/info/whatis#whatis).

By a loose definition, studies that examine one group of people before and after treatment could be considered clinical trials (Quest, 2001). However, by the strictest definition, a clinical trial compares a group of people receiving the experimental treatment (i.e., the treatment group) to a similar group of people who do not receive the treatment (i.e., the control group (Quest, 2001). Usually the control group is given an inert substance (i.e., placebo) so that any expectations participants may have about the experimental treatment will be the same in both groups and therefore theoretically will not influence the results (Quest, 2001). When the trial is finished, health differences between the two groups can be attributed to the treatment being tried instead of other factors, like the natural course of the disease, positive expectations of the drug's effects, age or gender (Quest, 2001).

The major difference between clinical trials and epidemiological studies (e.g., cohort or case-control study) is that in clinical trials, the investigators manipulate the administration of a new intervention and measure the effect of that manipulation. In contrast, epidemiological studies only observe associations (correlations) between the treatments experienced by participants and their health status or diseases (http://en.wikipedia.org/wiki/Clinical_trial).

Clinical trials may take on various forms. For instance, they may be randomized or not, placebo controlled or not, a crossover or parallel design, or multi-centered or of an experimental design layout (White, Ashby, & Brown, 2000). Although not a complete listing, the following sections examine various types of clinical trials one may come across in the medical literature.

3.1 RANDOMIZED CONTROLLED TRIAL (RCT)

In Randomized Controlled Trials (RCT) participants are assigned by chance to separate groups that compare different treatments. Neither the researchers nor the participants can choose which group. Using chance to assign people to groups means that the groups will be similar and that the treatments they receive can be compared objectively. At the time of the trial, it is not known which treatment is best. Both groups are followed up for a specified period and then groups are analyzed in terms of outcome defined at the outset. If the groups are similar at the outset, any difference should be due to the intervention.

The purpose of an RCT is to study interventions by objectively and fairly comparing their effect in similar groups. To achieve this, it is important to take steps to guard against bias (MRC, 2003). Failure to randomize, lack of blinding, and differential exclusion of people from the arms of a study can all bias results, typically leading to over-estimation of treatment effects. Randomization by a third party prevents the allocation of patients to treatment being consciously or subconsciously skewed. It reduces the risk that systematic differences between patients at baseline (i.e., confounders) will bias the result. In large trials, measurable prognostic variables are equally distributed between groups by chance. However, in small trials (e.g., less than 200 patients) minimization can be used as further "insurance policy". Here the randomization of new patients is weighted according to how pre-specified prognostic variables have distributed themselves in previous patients (MRC, 2003). This maximizes the chance that these factors will be equally distributed, while maintaining an element of randomization and the concealment of allocation necessary to ensure that unmeasured, but often equally biasing factors are not systematically skewed (MRC, 2003).

A report of a randomized controlled trial (RCT) should convey to the reader, in a transparent manner, why the study was undertaken and how it was conducted and analyzed (Moher, Schulz, & Altman, 2001). Despite several decades of educational efforts, RCTs still are not being reported adequately (Moher, Schulz, & Altman, 2001). Inadequate reporting makes the interpretation of RCT results difficult if not impossible. In response to this, in the mid-1990s, two independent initiatives to improve the quality of reports of RCTs led to the publication of the CONSORT (Consolidated Standards of Reporting Trials; Moher, Schulz, & Altman, 2001). CONSORT encourages transparency with reporting of the methods and results so that reports of RCTs can be interpreted both readily and accurately. The use of CONSORT seems to reduce inadequate reporting of RCTs and positively influence the manner in which RCTs are conducted (Moher, Schulz, & Altman, 2001).

All other clinical trials described below may or may not use the process of randomization.

3.2 PRECLINICAL TRIALS

Preclinical trials are those trials concerned with toxicity testing, pharmacokinetics or pharmacodynamics work (PK-PD), bioassays, determination of a dose that is both effective and safe, and bioequivalence studies. All these types of designs occur prior to a drug being administered to patients in a routine manner (White, Ashby, & Brown, 2000).

To prove that the compound works as is hypothesized and does not reproduce any negative side-effects, it is first thoroughly tested in animals (i.e., mice, rats, dogs, and monkeys). The purpose of this stage is to prove that the drug is not carcinogenic, and to understand how the drug is absorbed and excreted. Once a pharmaceutical company proves that the compound appears to be safe, and possibly effective in animals, the company will provide this information to the appropriate authority (Food and Drug Administration; FDA in the U.S. and Health Canada in Canada), requesting approval to begin testing the compound (experimental drug) in humans.

Scientific papers that use preclinical study designs may not be directly relevant to health technology assessment, but are certainly building blocks towards it (White, Ashby, & Brown, 2000). The drugs and treatments that successfully pass this stage in these studies then become the health technologies to be assessed in routine use (White, Ashby, & Brown, 2000).

3.3 CROSSOVER TRIALS

Crossover trials are often used to assess the effectiveness of a new drug when the disease being studied is chronic and its symptoms can be adequately controlled by medication but worsen when medication is withdrawn. It is a method of comparing two or more treatments or interventions in which subjects or patients, on completion of the course of one treatment, are switched to another. Typically, allocation to the first treatment is by random process. Participants' performance in one period is used to judge their performance in others, usually reducing variability (www.research-nurses.com/methodology_terminology.html. This crossover is done to address ethical concerns about depriving one group of a possibly beneficial treatment for the duration of the trial. Crossover trial designs encourage trial participation by promising all participants access to the experimental treatment (Quest, 2001).

One of the advantages of this design is that a smaller sample size is required; the within-subject variability is minimized by the subjects acting as their own control (White, Ashby, & Brown, 2000).

3.4 MULTI-CENTERE AND INTERNATIONAL CLINICAL TRIALS

A clinical trial that is conducted at more than one medical center or clinic is considered a multi-center research trial. Most large clinical trials are conducted at several clinical research centers. The benefits of multi-center trials include a larger number of participants, different geographic locations, various ethnic groups, the ability to compare results among centers, and thus increased generalizability of the study (www.answers.com/topic/**clinical**-trial-1).

Multi-center clinical trials are now commonplace, and reflect not only the need for increased numbers of research subjects but also the multidisciplinary nature of contemporary human research. Harmonization of ethical standards around the world is important given that we are in an environment of international research and medicine. The increasing number of international multi-center trials demands a uniformly high ethical standard for the conduct of research as does the ever-increasing technology and innovation in clinical practice (Tuffin & Chalmers, 1998).

3.5 EQUIVALENCE TRIALS

The aim of an equivalence trial is to show the therapeutic equivalence of two treatments, usually a new drug under development and an existing drug for the same disease used as standard active comparator (Wojdyla, 2005). In equivalence trials the null hypothesis states that the treatments are not equivalent, and the alternative hypothesis states that the treatments are equivalent (White, Ashby, & Brown, 2000). The finding in a trial that two treatments are equivalent does not require that both treatments were effective. It is equally compatible with the alternative hypothesis that neither was. In equivalence trials it is important to have means of confirming that both treatments were indeed effective (Wojdyla, 2005). The degree of certainty can be increased only by paying careful attention to the design of the equivalence trial, by being strict about matters of conduct, and by making additional checks during analysis. The most difficult issue relating to the analysis of an equivalence trial concerns which patient and which data from these patients to include (Wojdyla, 2005).

3.6 SCREENING TRIALS

Screening programmes are used mainly to detect a disease in its early stages when no symptoms are apparent. A large number of people are screened and those showing the early signs of the disease being screened for are then referred for further tests and contact with a specialist. By diagnosing a patient early, treatment of the disease is considered more effective. A screening programme is only worthwhile if treatment at the stage at which the disease is detected by screening means that survival after diagnosis is longer than it would have been without a screening programme. RCTs (i.e., screening trials) are used to evaluate the potential benefits of a screening programme (White, Ashby, & Brown, 2000).

Most screening tests have been developed to be noninvasive or mildly invasive (e.g., breast self-exams, mammograms, and pelvic exams). Screening tests exist for many of the more common cancers such as prostate cancer, breast cancer, colon cancer, lung cancer, and cervical cancer. Each screening test has an advisable age to begin screening and a recommended frequency at which the test should be performed.

3.7 SAFETY TRIALS

Safety trials are conducted on a large scale on drugs pre- and post-marketing. Those that are pre-marketing trials make up part of the evidence submitted to the regulatory authorities as to the efficacy and safety of a new drug. Post-marketing studies are important in order to pick up adverse events that are rare or occur in long-term use of a drug or in a specific patient population (White, Ashby, & Brown, 2000). An area that is related to HTA is the safety of equipment in hospital, for example magnetic resonance imaging machines or substances injected that show up on imaging machinery. Many of these substances are radioactive and so their safety has to be assured (White, Ashby, & Brown, 2000).

3.8 EXPLANATORY OR EFFICACY TRIALS

Explanatory trials (also known as efficacy trials) determine whether an intervention produces the expected result under ideal circumstances. They are frequently conducted in large tertiary-care, referral settings, which tend to have more specialized clinicians and better technical equipment than primary care facilities. Subjects in such studies typically live in areas with ready access to health centers and have accepted such referrals (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006).

Efficacy studies, especially phase III clinical trials, commonly use objective or subjective outcomes (e.g., symptom scores, laboratory data, or time to disease recurrence) to determine intermediate outcomes (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). External validity is limited if study protocols do not reflect clinical practice. Efficacy trials are required for approval purposes, and investigators design study durations and treatment modalities to demonstrate an effect and ensure safety. Such trials may not last as long as therapy would in everyday practice (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). Additionally, they may rely on strict diagnostic criteria that are usually not employed in primary care settings. Investigators need to ensure or measure compliance to determine whether an intervention works (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006).

3.9 PRAGMATIC OR EFFECTIVENESS TRIALS

Pragmatic trials (also known as effectiveness trials) measure the degree of beneficial effect under "real world" clinical settings. Hence, hypotheses and study designs on an effectiveness trial are formulated based on conditions of routine clinical practice and on outcomes essential for clinical decisions (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). For effectiveness trials, settings should reflect the initial care facilities available to a diverse population with the condition of interest. For persons with rare or severe diseases or those requiring high-risk interventions, such as organ transplantations, specialized secondary or tertiary care settings may provide initial care (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006).

In pragmatic trials, eligibility criteria must allow the source population to reflect the heterogeneity of external populations: the full spectrum of the human population, their co-morbidities, variable compliance rates, and use of other medications or therapies (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). Health outcomes, relevant to the condition of interest, should be the principal outcome measures in effectiveness studies. Intermediate outcomes are adequate only if empirical evidence verifies that the effect of the intervention on an intermediate endpoint predicts and fully captures the net effect on a health outcome (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006).

For pragmatic trials, study durations should mimic a minimum length of treatment in a clinical setting to allow the assessment of health outcomes (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006). Treatment modalities should reflect clinical relevance and diagnosis should rely on diagnostic standards that practicing physicians use. Investigators should define compliance as an outcome measure because unpredictable or poor compliance can render an efficacious treatment ineffective (Gartlehner, Hansen, Nissman, Lohr, & Carey, 2006).

3.10 BLINDING OR MASKING TRIALS

Blinding is a design feature that keeps exposure status or disease status secret from a design feature, that keeps exposure status or disease status secret from at least one set of study participants. Single blind trials are those in which the patient (subject) or the investigator (exposure disease evaluator) is unaware of treatment group assignment or case-control status. In a double blind trial, both the clinician (investigator/evaluator) and patient are unaware of the treatment group or case-control status. Double blinding, or masking both the patient and the provider or observer of that intervention has been shown to have a strong effect on treatment outcome. Ideally, neither the researcher-observer, treating clinician, patient, nor the statistician should know which treatment group a person was assigned (http://www.nsc.nhs.uk/glossary/glossary_ind.htm). Double-blind designs lend themselves well to drug studies in which identically packaged active drugs or substances are compared to inactive or placebo ingredients. In a triple blind trial, in addition to the patient and investigator being unaware of the group assignment or case-control status, the data safety monitoring committee or the statistician is also left unaware.

Blinding or masking involves keeping secret group assignment (e.g., to treatment or control) from the study participants or investigators. Blinding is used to protect against the possibility that knowledge of assignment may affect patient response to treatment, provider behaviours (i.e., performance bias) or outcome assessment (i.e., detection bias). However, blinding is not always practical (e.g., when comparing surgery to drug treatment). The importance of blinding depends on how objective the outcome measure is; blinding is more important for less objective outcome measures such as pain or quality of life (http://www.nsc.nhs.uk/glossary/glossary_ind.htm).

3.11 PLACEBO-CONTROLLED

A method of investigation in which an inactive substance (i.e., placebo) is given to one group of participants, while the treatment or intervention being tested is given to another group. The results obtained in the two groups are then compared to see if the investigational treatment is more effective in treating the condition. An ethical concern is that use of placebos may deny patients potentially helpful treatment. A practical concern is that placebo-controlled trials in this setting may be of little interest or value to patients or investigators and that only a comparison of the new treatment with existing treatment will provide useful data (Ellenberg & Temple, 2000).

Placebo controls are commonly used in clinical trials of investigational treatments because they have important advantages. In recent years, some have criticized the use of placebo-controlled trials when effective alternative therapy exists, regardless of the expected effect of the therapy (Ellenberg & Temple, 2000).

4.0 PURPOSE AND STRUCTURE OF CLINICAL TRIALS AND ETHICS

The purpose of clinical trials is to study interventions by objectively and fairly comparing their effects in similar groups. A clinical trial is only done when there is some reason to believe that the treatment being studied may be valuable to the patient. Treatments used in clinical trials are often found to have real benefits.

Researchers conduct studies of new treatments to answer the following questions:

- Is the treatment helpful?
- How does this new type of treatment work?
- Does it work better than other treatments already available?
- What side effects does the treatment cause?
- Are the side effects greater or less than the standard treatment?
- Do the benefits outweigh the side effects?
- In which patients is the treatment most likely to be helpful?

From an ethical perspective, clinical studies are "human experiments", and thus need to: (1) address a legitimate research question, (2) ensure the patient is informed and willing to participate in the trial, and (3) give the patient the opportunity to decline entry or withdraw from the trial at any stage.

The clinical testing of experimental drugs in humans is normally done in four phases, with more people included in each subsequent phase. Before moving to the next phase of development the data are carefully analyzed to ensure the experimental drug is at least safe and well tolerated. In North America, after successful completion of Phase I to III testing, a company submits the results of all of the studies to the FDA or Health Canada to obtain a New Drug Approval (NDA). Once the regulatory authority grants a company with a NDA, the company can market the drug (i.e., medication) to the public. Then, additional testing (post-marketing or phase IV) to look at the long-term safety continues (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html).

In the proceeding section we present an overview of each of these four phases.

4.1 PHASE I STUDY: BASIC PHARMACOLOGICAL AND TOXICOLOGY INFORMATION

In Phase I clinical trials the main goal is to obtain basic pharmacological and toxicology information. Researchers test a new drug or treatment in a small group of healthy human volunteers (20-80) for the first time to: (1) evaluate its safety, (2) determine a safe dosage range, and (3) identify side effects. Typically trials are conducted in a hospital setting where they can be monitored and treated in the event of any side effects. Volunteers are usually paid for their participation and for the most part tend to be men (approximately 30 years of age on average; http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html).

The purpose of these studies is to determine how the experimental drug is absorbed, metabolized, and excreted in humans. Additionally, they seek to determine what types of side effects occur as the dosage of the drug is increased. Any beneficial effects of the drug are also noted (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html).

4.2 PHASE II STUDY: IDENTIFY DOSE RANGE OF A DRUG

In Phase II clinical trials, the goal is to identify the dose range of a particular drug. The study drug or treatment is given to a larger group of people (40-100) to: (1) see if it is effective and (2) further evaluate its safety.

Once an experimental drug has been proven to be safe and well tolerated in healthy volunteers, it must be tested in the patients that have the disease or condition that the experimental drug is expected to improve or cure. In addition to ensuring that the experimental drug is safe and effective in the patient population of interest, Phase II studies are also designed to evaluate the effectiveness of the drug (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html). The second phase of testing may last from several months to a few years and may involve up to several hundred patients.

Most Phase II studies are well-controlled, randomized trials (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html). Thus, one group of patients or subjects receives the experimental drug, while a second "control" group receives a standard treatment or placebo. Placement of the subject into the drug treatment or placebo group is by random chance (as if by the flip of a coin). Often these studies are "double-blinded" (i.e., neither the patient nor the researchers know who is getting the experimental drug).

Additionally, Phase II studies are often designed to determine the dosage with the least number of side effects that is most effective (i.e., correct dosage). These are often referred to as dose-ranging studies. In general, the purpose of Phase II studies is to provide the pharmaceutical company and the FDA or Health Canada with comparative information about the relative safety of the experimental drug, the proper dosage needed to treat the condition, and the drug's effectiveness (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL% 20TRIAL.html). Only about one-third of experimental drugs successfully complete both Phase I and Phase II testing (http://www.huntington-study-group.org/WHAT%20IS%20A%20 CLINICAL%20TRIAL.html).

4.3 PHASE III STUDY: COMPARE EFFECTS OF DIFFERENT TREATMENTS

In Phase III studies, the study drug or treatment is given to large groups of people (more than 200), with the disease or condition of interest, and usually from different institutions in several countries. The objective in this phase is to: (1) further determine the treatment's effectiveness, (2) monitor side effects, (3) compare it to commonly used treatments, and (4) collect information that will allow the drug or treatment to be used safely (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html). Most Phase III studies continue to be randomized and blinded. Outcome measures usually include survival, disease-free survival, response and toxicity.

The large-scale testing provides the pharmaceutical company as well as the FDA or Health Canada with a more thorough understanding of the drug's effectiveness, benefits and risks, and range and severity of possible adverse side effects. Phase III studies typically last several years. Seventy to 90 percent of drugs that enter Phase III studies successfully complete this phase of testing (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html).

4.4 PHASE 4 STUDY: POST-MARKETING

After successful completion of Phase I to III testing, a company submits the results of all of the studies to the FDA or Health Canada to obtain a New Drug Application (NDA). Once the regulatory authority grants a company with a NDA, the company can market the drug (medication) to the public. Phase IV studies are typically carried out after the drug or treatment has been marketed. These studies continue testing the study drug or treatment to collect information about their effect in various populations, and any side-effects associated with long-term use. This additional testing is also known as post-marketing (http://www.huntington-study-group.org/WHAT%20IS%20A%20CLINICAL%20TRIAL.html).

5.0 EVALUATION OF RANDOMIZED CONTROLLED TRIALS

5.1 ADVANTAGES AND DISADVANTAGES OF RCTS

Although the RCT has theoretical advantages over other study designs, McCulloch, Taylor, Sasako, Lovett, and Griffin (2000), have argued that experimental studies comparing treatment effect estimates in randomized and non-randomized studies have not consistently confirmed this. As a result, they caution that the superiority of RCTs should not be accepted as self-evident. Still, among the main advantages of RCTs cited are the following: (1) RCTs provide a rigorous evaluation of a single variable in a defined patient group; (2) they potentially eradicate bias by comparing two or more identical groups; and (3) they allow for a meta-analysis to be conducted.

The main disadvantages of RCTs are that they: (1) are expensive and time-consuming; (2) often include too few patients or too short a follow-up period; (3) involve surrogate endpoints that are often used in preference to clinical outcome measures; (4) employ imperfect randomization; (5) often do not randomize all eligible patients; and (6) typically fail to blind assessors to the randomization status of patients. According to Black (1996), the limitations of randomized trials can be seen as deriving from either the inherent nature of the method (a limitation in principle) or from the way trials are conducted (a limitation in procedure). The importance of this distinction is that while little can be done about the former, improvements in the conduct of randomized trials could resolve some or all of the latter.

5.2 SPECIFIC OBSTACLES TO RANDOMIZED CONTROLLED TRIALS (IN SURGERY)

An analysis of general surgical work in a large UK hospital showed that only 24% of the treatments used were based on RCT evidence, compared with over 50% for inpatient general medicine (McCulloch, 1999). A recent analysis of the illnesses and treatments most commonly encountered in general surgery suggested that less than 40% of operative treatments were amenable to study using an RCT design (McCulloch, 1999).

Historically, the majority of common general surgical operations were introduced before 1920, long before the importance of the randomized controlled trial was appreciated anywhere in the medical profession (McCulloch, 1999). It is always hard to do randomized trials of well-established treatments, because the attachment of both doctors and patients to the familiar prevents the level of open-minded doubt necessary to achieve "equipoise". For this reason, many operations, together with time-honored medical treatments like morphine have largely escaped the rigors of the RCT versus placebo for their original indications (McCulloch, 1999). Moreover, ethical considerations have been considered barriers to the use of placebo-controlled investigations for surgical procedures (Kent, 2002).

The nature of treatment by surgical operation provides at least two good reasons for not performing RCTs: (1) many of the conditions treated by surgery are of a mechanical nature, and (2) in some cases (such as relief of mechanical bowel obstruction) the superiority of the mechanical solution offered by operation over non-treatment is self-evident. In many surgical scenarios, the benefits are so clear that no one would consider a trial ethical or remotely sensible. Thus, there is no question of a placebo-controlled trial of repair of inguinal hernia, relief of mechanical bowel obstruction, or drainage of abscesses (McCulloch, 1999).

Additionally, surgery is a skilled, multi-step process, and this makes RCT designs difficult to deliver in surgical studies for two reasons. First, there is a learning process in every new operation, even for a fully trained surgeon unfamiliar with the particular procedure. Serious bias can be easily introduced if this is not acknowledged and measured or eliminated, especially for trials of new versus older procedures. Second, there is inherent variation in the way the procedure is performed by every individual, and this cannot be eliminated. Surgeons stress the need for quality control in the technical aspects of any procedure under trial, but are acutely aware of the difficulty of the task (McCulloch, 1999). We know that large variations in outcome are observed between surgeons performing similar operations in the same population.

The quality and quantity of randomized trials of surgical techniques is acknowledged to be limited (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Some aspects of surgery present special difficulties for randomized trials. The RCT has theoretical advantages over other study designs, but experimental studies comparing treatment effect estimates in randomized and non-randomized studies have not consistently confirmed this (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Thus, the superiority of RCTs should not be accepted as self-evident. Until recently, most studies of operations were retrospective case series, with RCTs accounting for less than 10% of the total. Treatments in general surgery are half as likely to be based on RCT evidence as treatments in internal medicine. (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

As described by McCulloch et al., (2002) this raises the important question of "why then, is surgery so deficient?" (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002)

5.2.1 HISTORY

A comprehensive review of the evidence base is needed to indicate areas warranting new trials of old techniques (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). History did not favor the validation of surgery by RCTs. After the invention of anaesthesia and antiseptic techniques, surgical treatments were rapidly developed for many previously untreatable conditions. Many current operations were therefore introduced well before

randomized trials became established in medicine, unlike most modern drugs (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Once a treatment is accepted as standard, testing it against placebo becomes difficult. For fields such as cardiac surgery, transplantation, orthopaedics, and neurosurgery, however, which have developed rapidly since 1950, surgeons cannot fall back on history to explain the lack of rigor in surgical research (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.2 COMMERCIAL COMPETITION AND PRESTIGE

According to McCulloch, Taylor, Sasako, Lovett, and Griffin (2002), doctors can be tempted to ignore evidence that threatens their personal interests. Objectivity about procedures central to a surgeon's reputation is difficult, and RCTs may seem threatening. Private sector competition may affect surgeons particularly strongly, and it arguably influenced the introduction of laparoscopic cholecystectomy (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). A consensus conference in 1994 quoted many reports of increased bile duct injuries and only two RCTs. The benefits shown were not overwhelming against the evidence of possible harm, but further RCTs were declared infeasible because the technique was already so widespread. Surgeons' eagerness to learn the operation seemed related more to commercial concerns than to concern for patients (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.3 SURGEONS' EQUIPOISE

Career surgeons are selected for traits that include comfort with making important clinical decisions quickly with incomplete information (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). This quality, required for decisive action during operations, may make it difficult for them to be consciously uncertain which of the two treatments is better. This state of equipoise, however, is a prerequisite for performing RCTs (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). If confirmed, surgeons' equipoise may need to be accommodated by including parallel, non-randomized, preference arms alongside RCTs (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.4 LACK OF FUNDING, INFRASTRUCTURE, AND EXPERIENCE OF DATA COLLECTION

These are real and major problems for surgical trials. The difficulty is partly self-inflicted as funding bodies are influenced by the poor quality of much previous surgical research (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). What is required is a change to a culture of cooperation rather than one of competition. Such a change would in turn, facilitate the creation of large groups to perform specific trials, thereby attracting funding and developing the infrastructure. Naturally, this change would require support from the bodies responsible for funding clinical research (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.5 LACK OF EDUCATION IN CLINICAL EPIDEMIOLOGY

McCulloch, Taylor, sasako, Lovett, and Griffin (2002) argued that subjectively, surgeons' knowledge of clinical epidemiology remains poor despite relevant publications in surgical journals. These authors stated that there is no objective evidence that surgeons receive less specific education than other groups of doctors. Still, surgeons recruit patients for cancer chemotherapy trials less readily for trials of surgical technique (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). What remains unclear is whether there really is a lack of education in clinical epidemiology among surgeons. Thus, this requires investigation, and if it is demonstrated that lack of education is indeed a real occurrence, then the bodies responsible for postgraduate surgical education and training must move to correct this (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.6 RARE CONDITIONS AND LIFE THREATENING AND URGENT SITUATIONS

Emergency surgery often occurs outside normal working hours and involves urgent lifesaving treatment, thereby making consent and randomization difficult. Moreover, uncommon conditions are difficult to investigate when accrual of patients takes over two years (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Rare conditions and life threatening, urgent situations will always be challenging areas for RCTs in surgery. However, they have been successfully studied in other disciplines. For instance, paediatric oncologists have illustrated the enormous value of cooperation through their success in trials on childhood leukaemia (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.7 THE LEARNING CURVE

Some authors suggest that RCTs of new operations should begin with the first patient. Operations, however, are complex procedures, and quality in performance requires frequent repetition over time. Learning curves of similar lengths are reported for disparate operations (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). During the learning curve, errors and adverse outcomes are more likely. Randomizing between a familiar and an unfamiliar operation therefore introduces bias against the latter. This problem for surgical RCTs has few parallels in drug trials (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

The learning curve needs to be recognized and evaluated using appropriate statistical techniques. Moreover, the trial methodology could also be modified. Although not yet tested in practice, in theory, patients could be randomized to surgeons (not operations), who would perform their operation of preference (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.8 DEFINITION OF INTERVENTION AND QUALITY CONTROL MONITORING

The technical quality of operations undoubtedly affects outcome. Poor quality surgery represents failure to deliver the intended treatment, causing a difference between efficacy and effectiveness (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Trials then measure deliverability, not efficacy. Quality control failures may narrow important differences in the surgery received and may influence outcomes. Defining and enforcing minimum quality standards may be difficult for surgical trials (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

Variations on an operation are common and may influence success rates. When comparing operations, clear definitions are therefore needed of the limits on acceptable technical variation. A standard description may be necessary, proscribing all modifications. Unlike in drug trials where treatments are usually simple to define in exact terms, in surgery imprecise definitions may result in the delivery of overlapping treatments (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

Precisely defined photographic or video evidence and/or pathological specimens could document the nature and quality of the treatment delivered. Norms for pre-trial success rates and complications could provide a basis for defining acceptable quality, making reliable surgical audit data essential for participation in RCTs (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.9 DEVELOPMENT VERSUS RESEARCH

RCTs consume substantial resources and are therefore not justified for some questions about small modifications to treatments. Surgical technique typically progresses via such modifications, which individually are unlikely to produce detectable benefits, but which collectively may do so (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). For instance, during the historical progression through hand washing via the use of antiseptics to the aseptic surgical environment, the change in morbidity from surgical infection was huge, but the increment with each step was small enough to allow persistent skepticism (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Small randomized trials of components of this progression showed no benefit. If a positive RCT were required before adopting each small improvement, most would be rejected, and progress would be slowed. RCTs are appropriate where a clear, clinically important choice exists between contrasting alternatives. For smaller changes, an industrial paradigm may be needed (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Surgeons should adopt industrial quality assessment techniques to evaluate changes in technique where RCTs are inappropriate (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Sequential approaches such as CUSUM and the "control curve" are also applicable to surgical innovation (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.10 PATIENTS' EQUIPOISE

Three types of RCT are commonly described as "surgical". **Type 1 trials** are standard RCTs comparing medical treatments in surgical patients and they account for 75% of "surgical trials" (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). **Type 2 trials** involve comparing surgical techniques, but pose the problems described above (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). **Type 3 trials** compare surgical and non-surgical treatments; these trials pose particular difficulties with the equipoise of patients. Patients often reject RCTs because they do not wish their treatment to be decided by chance (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Type 3 trials increase this discomfort because the adverse effects of the options often differ enormously and the surgical option is irreversible. Eighty two percent of problems preventing type 3 trials are related to patients' equipoise (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Examples of choices include aspirin versus carotid endarterectomy to prevent embolic stroke and goserelin versus castration for prostate cancer. Such trials may recruit slowly, or select an unusual subgroup of patients, making them impractical or their results difficult to generalize (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Resolution to type 3 trials may derive from decision analysis techniques and carefully designed composite end points to reflect the contrasting possible outcomes of trial arms (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

5.2.11 BLINDING

It has been demonstrated that of those examined, only a third of surgical trials had adequate blinding of patients and/or surgeons (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Although blinding is particularly difficult in surgical trials, creative solutions such as the use of standardized wound dressings can result in success (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

6.0 PROPOSED FRAMEWORK FOR CLINICAL RESEARCH (IN SURGERY)

The trade-offs between the internal validity of RCTs and their applicability in standard practice is a recognized problem in HTA. RCTs can be impractical for research on important health problems (e.g., diagnosis and treatment of traumatic injury and intervention in the intensive care setting), as well as for rare events (Heitman, 1998). In such contexts, it may be difficult to distinguish between innovative treatment and experimentation, and even more difficult to determine when to undertake a formal trial of a new intervention developed in the course of patient care (Heitman, 1998). Because of the inherent difficulty of obtaining consent to participation, accruing a statistically significant subject pool can take so long that the technology under study may be outdated or universally accepted by the end of a protocol (Heitman, 1998). Still, it is no longer possible to appeal to the mystique of professional expertise, when asked to justify our decisions or our results. The less secure the evidence base for our practice, the less likely it is to be able to withstand pressure from public and political voices. Surgeons, like other doctors, need evidence-based medicine, because the alternative is policy-based medicine (McCulloch, 1999). There are ways to address the practical difficulties peculiar to surgery, and they need to be taken seriously and instituted (McCulloch, 1999). According to McCulloch, Taylor, Sasako, Lovett, and Griffin (2002), there is a need for a framework that reflects the difficulties of evaluation in surgery.

6.1 AUDIT DATA COLLECTION

The baseline for the scientific study of surgery is routine collection of comprehensive data about practice and outcomes. The culture and organization necessary for this should permit easy participation in trials, whereas where these are absent, trialists have to develop the trial infrastructure and run it simultaneously (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Surgeons need the resources to record a meaningful audit dataset, entailing considerable investment in data acquisition and management resources (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

6.2 CONTINUOUS PERFORMANCE EVALUATION

Systems for continuous quality control, using instruments such as CUSUM, CRAM, or VLAD plots or control curves should be used for the analysis of technical innovations. Indications of outcome changes from this surveillance should lead to an audit, using decision analysis techniques to determine whether an RCT is warranted (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Where it is not, continuing prospective data collection and regular re-evaluation using Bayesian analysis provides the best available data on outcome changes and allows reconsideration of the need for an RCT (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

6.3 CONDUCT OF RANDOMIZED CONTROLLED TRIALS (RCTS)

When RCTs are necessary, they should routinely be preceded by preliminary phase 2S (phase 2 surgical) studies. These would develop satisfactory definition criteria for the procedure, test measures of surgical quality, define suitable end points, estimate the required sample size, and analyse the learning curve of participants (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Such studies would reduce the problems of timing surgical RCTs, and randomization could be introduced early using "tracker" designs if desired. During randomized data entry, continuous quality control should be linked to preplanned interim analyses by the trial review committee and appropriate stopping rules (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Objective validation of quality should evaluate images, pathological specimens, and outcome data against criteria drawn up in the phase 2S study. Parallel preference arms may be used to improve overall power and evaluate generalizability (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). For type 3 trials, end point design and decision analysis tools to help patients understand their choices may be important (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002).

6.4 OTHER SOURCES OF EVIDENCE

As we have seen, historically, the surgical literature is poor in RCTs. Meta-analysis of non-randomized evidence should therefore be used wherever appropriate (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). Where RCTs are difficult for sound reasons, prospective non-randomized designs that minimize known biases should be considered sympathetically by journals and funding bodies (McCulloch, Taylor, Sasako, Lovett, & Griffin, 2002). [For a more detailed discussion of non-randomized or non-experimental design studies, please refer to Module 5 entitled "How to assess observational, outcome and audit research in surgery" in our Health Technology Assessment Module and Workshop Series.]

7.0 CONCLUDING SUMMARY

The primary influence of research publications is to inform decisions by patients, practitioners, payers, and policy makers (Bloom, Retbi, Dahan, & Johnsson, 2000). Inadequate studies have no value, since they may lead to overuse of ineffective and underuse of effective technology by clinicians (Bloom, Retbi, Dahan, & Hohnsson, 2000).

In health technology assessment the question: "Does the technology work?" is most easily answered using standard statistical methods (White, Ashby, & Brown, 2000). As HTA develops, it is likely to need more complex studies and methods to answer questions about packages of interventions and interactions between them (White, Ashby, & Brown, 2000). This poses a specific challenge of bridging the gap between current medical methodological and statistical training, and a health technology assessment perspective (White, Ashby, & Brown, 2000).

In this module we reviewed the types of clinical studies that form important sources of clinical evidence. The RCT is accepted as the best available method to prove cause and effect while minimizing bias in observed (patient) and observer (practitioner, evaluator). However, not everyone is in agreement that the RCT is the best method available to collect and evaluate biomedical evidence (Bloom, Retbi, Dahan & Johnsson, 2000).

As we have seen, there are limitations to conducting randomized controlled trials (RCTs). Two important things to consider are: (1) the scale and pace of health technology developments is greater than the capacity to fund and carry out RCTs, and (2) RCTs may occasionally be inappropriate, impossible, or inadequate for HTA (Stevens, Raftery, & Roderick (2005).

Considering the limitations of relying on experimental (specifically, clinical trials) studies for health technology assessment, it seems necessary to supplement our discussion with other sources of information. Non-experimental designs may provide a potential alternative or complement to RCTs. For an extended discussion of these designs, refer to Module 5 of our Health Technology Assessment Modules and Workshop series.

7.1 REVIEW OF MODULE OBJECTIVES

By the end of this module, participants should be able to:

(1) Describe some of the types of clinical trials appearing in the medical literature. Specifically: Randomized Controlled Trials, Preclinical Trials, Screening Trials, Crossover Trials, Multi-center and International Clinical Trials, Equivalence Trials, Screening Trials, Safety Trials, Explanatory or Efficacy Trials, Pragmatic or Effectiveness Trials, Blinding or Masking Trials, and Placebo-Controlled Trials).

(2) Identify the purpose and structure of clinical trials.

(3) Discuss the advantages and disadvantages to conducting clinical trials.

(4) Understand the obstacles to conducting clinical trials.

(5) Propose a framework for clinical research in surgery.

8.0 REFERENCES

Agence d'evaluation des technologies et des modes d'intervention en sante' (AETMIS) & International Society of Technology Assessment in Health Care (ISTAHC). (2000 – 2001). Health Technology Assessment: Decision-Making for Health. Distance Learning Course. Units 1 to 12. Course content developed by a team of researchers and instructors at McGill University, University of Montreal and AETMIS in Montreal, Canada. http://aetmis.gouv.qc.ca/e-learning/mod/forum/discuss.php=73

Alberta Heritage Foundation for Medical Research; AHFMR. (2003). *Health technology assessment: Alberta's needs and future directions. Summary of results of a consultation workshop.* Edmonton, AB. Canada.

American Medical Association. (1992). Evidence-based medicine: A new approach to teaching the practice of medicine. Based on the *Users' Guides to Evidence-based Medicine and reproduced with permission from JAMA. 268 (17)*: 2420-5.

Australian Government. (1996). "Measuring and improving outcomes: A note from the editor." *Better Health Outcomes Newsletter, 2(1)*. Department of Health and Aging. http://www7.health.gov.au/pq/bho/1996/v2n1/outcomes.htm#definitions

Black, N. (1996). Why do we need observational studies to evaluate the effectiveness of health care. *British Medical Journal, 312*: 1215-1218.

Bloom, B.S., Retbi, A., Dahan, S., & Johnsson, E. (2000). Evaluation of randomized controlled trials on complementary and alternative medicine. *International Journal of Technology Assessment in Health Care, 16(1)*; 13 –21.

Danish Institute for Health Technology Assessment; DIHTA. (2001). Health Technology Assessment Handbook. 1st Edition. Kristensen, F.B., Horder, M., & Poulsen, P.B. (Eds.). National Board of Health.

Ellenberg, S.S. & Temple, R. (2000). Placebo-controlled trials and active-control trials in the evaluation of new treatments. Part 2: Practical issues and specific cases. *Annals of Internal Medicine, 133(6)*; 464-470.

First International Conference on Health Promotion Ottawa. (1986). Ottawa Charter for Health Promotion. WHO/HPR/HEP/95.1. http://www.who.int/hpr/NPH/docs/ottawa_charter_hp.pdf

Flaherty, R.J. (2004). A simple method for evaluating clinical literature. *Family Practice Management*. http: www.aafp.org/fpm/20040500/47asim.html

Gartlehner, G., Hansen, R.A., Nissman, D., Lohr, K.N., & Carey, T.S. (2006). Criteria for distinguishing effectiveness trials in systematic reviews. *Technical Review, 12*. Prepared for: Agency for Healthcare Research and Quality (AHRQ). AHRQ Publication No. 06-0046. April. http://ahrq.gov/downloads/pub/evidence/pdf/efftrials/efftrials.pdf

Greenhalgh, T. (1997). Assessing the methodological quality of published papers. *BMJ, 315*; 305-308.

Kent, C. (2002). Placebo surgery. http://www.worldchiropracticalliance.org/tcj/2002/sep/sep2002kent.htm.

Khan, K.S., ter Riet, G., Glanville, J., Sowden, A.J., & Kleijnen, J. (Eds.) (2001). Undertaking systematic reviews of research on effectiveness. CRD's Guidance for those Carrying out or Commissioning Reviews. CRD Report Number 4 (2nd Edition). NHS Centre for Reviews and Dissemination, University of York. York Publishing Services Ltd.

Lilford, R.J., Richardson, A., Stevens, A., Fitzpatrick, R., Edwards, S., Rock, F., & Hutton, J.L. (2001). Issues in methodological research: perspectives from researchers and commissioners. *Health Technology Assessment, 5(8)*; 1 – 62.
Macdonald, G. (2003). Using systematic reviews to improve social care. Social Care Institute for Excellence. SCIE Reports: No.4

McCulloch, P. (1999). Evidence-based surgery: do we need it, and can we get it? http://www.cebm.utoronto.ca/syllabi/gs/intro.htm. Liverpool, April 1999-04-13

McCulloch, P., Taylor, I., Sasako, M., Lovett, B., & Griffin, D. (2002). Randomised trials in surgery: Problems and possible solutions. *British Medical Journal, 324*; 1448-1451.

Medical Research Council (MRC). (2003). Health technology assessment in surgery. www.mrc.ac.uk Accessible at: http://www.mrc.ac.uk/pdf-health_tech_assessment.pdf

Moher, D., Schulz, K.F., & Altman, D.G. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *Annals of Internal Medicine, 134(8)*; 657-662.

Quest. (2001). Simply stated . . . clinical trials. *Quest, 8(5)*; 1-3. Accessed at: http://www.mdausa.org/publications/Quest/q85ss.cfm

Rational Pharmaceutical Management Plus Program. (2001). Drug and therapeutics Committee Training Course. Session 3: Assessing Drug Efficacy. Management Sciences for Health. E-mail: rpmplus@msh.org

Stevens, A.J., Raftery, J., & Roderick, P. (2005). Can health technologies be assessed using routine data? *International Journal of Technology Assessment in Health Care, 21:1*; 96 – 103.

Tuffin, S. & Chalmers, D. (1998). Impacts of using an evidence-based approach to health care decision-making: An ethical perspective. *Evidence-based Health Advice Workshop,* 4 – 5 November. The Menzies Foundation in association with the Health Advisory Committee of the NHMRC. Accessible at: http://www.menziesfoundation.org.au/conferences/ebm/ebm12atu.htm.

White, S.J., Ashby, D., & Brown, P.J. (2000). An introduction to statistical methods for health technology assessment. *Health Technology Assessment, 4(8)*; 1-59.

Wojdyla, D. (2005). Cluster randomized trials and equivalence trials. UNDP/UNFPA/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction. World Health Organization. Accessible at: http://www.gfmer.ch/Medical_education_En/PGC_RH_2005/pdf/Cluster_Randomized_Trials.pdf

World Health Report (WHR). (2000). "Chapter 1: Why do health systems matter?" www.who.int/whr/2000/en/whr00_ch1_en.pdf

9.0 APPENDICES

9.1 APPENDIX A: CLINICAL TRIALS INSIGHT

Study Phase

Four phases of clinical trials and drug development exist and are defined below.

Phase I
Purpose: Determine tolerability and dosage.

Initial safety trials on a new medicine, usually conducted in normal male volunteers. An attempt is made to establish the dosage range tolerated by volunteers for single and multiple dosages. As well as the drug's tolerability profile and dosage range, these studies also determine how a drug is absorbed, distributed, metabolised, and excreted, and the duration of its action.

- Patients are evaluated instead of volunteers in Phase I clinical trials:
    i. When the medicine is too toxic to test ethically in volunteers (e.g. many anticancer agents).
    ii. When the expected therapeutic ratio is too narrow to test the medicine ethically in volunteers (e.g. antiarrhythmics).
    iii. When it is believed that the therapeutic dose in patients will be greater than normal volunteers can tolerate (e.g. a neuroleptic).
    iv. To study patients with a serious disease who are in remission.
    v. To study patients who do not have the specific concomitant illness that the new medicine is designed to treat.
- Phase I clinical trials may enrol patients and volunteers - although in separate trials if it is believed that patients may metabolise the medicine differently than normal volunteers, such as in epileptic patient populations, because these patients are often receiving medicines that affect hepatic microsomal enzymes.
- Pharmacokinetic trials are usually considered Phase I trials regardless of when they are conducted during a drug's development.

Phase II
Purpose: Provide a measure of efficacy in addition to short-term tolerability.

Phase II studies are conducted in patients who have the disease or condition that the drug is intended to treat. Other Phase II study objectives include determining the minimum dose that is maximally effective, or that is sufficiently effective without undue toxicity.

When it is both possible and useful, Phase II studies should be controlled investigations involving a placebo or standard therapy comparator.

Early and late Phase II studies

Patients selected for early Phase II studies should ordinarily be free of haematological, hepatic, renal, cardiac, or other serious diseases.

Patients with concomitant diseases and therapy may be included in late Phase II studies, since they are representative of certain segments of the population that would receive the investigational drug if it gained marketing approval.

Phase III
Purpose: Confirm efficacy, monitor adverse reactions from long-term use.
In Phase III studies, a drug is tested under conditions more closely resembling those under which the drug would be used if approved for marketing. The goal is to gather additional information about efficacy and tolerability that is needed to evaluate the overall benefit-risk relationship of the drug and to provide an adequate basis for physician labeling.

Note

FDA approval/disapproval and pivotal studies.
FDA approval/disapproval decisions are based on the results of adequate and well-controlled (pivotal) studies. To be considered pivotal, a study must meet at least the following 4 criteria:

1. Be controlled using placebo or a standard therapy.
2. Have a double-blinded design when such a design is practical and ethical.
3. Be randomized.
4. Be of adequate size. Sample size calculation requires many assumptions about the results to be obtained with the treatment and population being studied. Because considerable clinical judgement is used in making these assumptions, faulty presumptions frequently result in studies of inadequate statistical power. Study sample size is a common clinical trial design flaw.

Phase IV
Purpose: Provide ongoing data after regulatory (e.g. FDA) approval.
Phase IV clinical trials are undertaken for reasons such as:

1. To satisfy a regulatory request (i.e. that Phase IV trials be conducted after approval).
2. To evaluate drug efficacy and tolerability under conditions of widespread use, including in patient groups not well represented in well-controlled Phase III trials.
3. For special-purpose testing, such as cost-benefit or cost-effectiveness studies, that attempt to, for example, find specific competitive advantages of using one product over another drug or therapy.

Almost any type of clinical study may be conducted during Phase IV, including postmarketing surveillance, pharmacoepidemiology, marketing-oriented, and clinically-oriented studies. Phase IV trials are often of a larger scale than are premarketing studies, and are sometimes less rigorously controlled.

Appendix - Study Design

Design	*Definition*
Case report	This term is used for therapeutic-use case reports. This term may also be used for drug-interaction case reports that do not have an adverse reaction (e.g. an interaction that alters the pharmacokinetic characteristics of one of 2 drugs given).
Case control	The traditional definition of a case-control study is that of an observational epidemiological study that starts with a group of individuals receiving a treatment/having a disease (cases) and a well-matched group of similar individuals not receiving treatment/not having a disease (controls). The relationship of a risk factor to the disease is evaluated by determining how frequently the risk factor is present in the cases and controls. In recent times the case control study has been seen as a variation on a cohort study; the controls are a representative sample of the study population (cohort) from which all the outcomes (cases) are identified. Case-control studies may be longitudinal or cross-sectional.
Cohort	A cohort study involves a group that is exposed, and followed prospectively to a point in time when subjects are evaluated for differences in the frequency of one or more outcomes from an unexposed group. Subjects from a defined group are not allocated to treatment but rather the investigator measures treatments that are already present. The groups being compared may differ by factors other than the treatment of interest. e.g. drug therapy in community-based menopausal Belgian women followed for 5 years (groups being compared may be no treatment, calcium, and hormone replacement therapy). Usually subjects and/or clinical management characteristics are linked with various health outcomes.Case-control, cross-sectional and prospective are all subtypes of the cohort study design.
Crossover	The subject group is subdivided, and receive all treatments in a parallel and reciprocal manner, usually separated by a washout period. Allocation to the first treatment is typically randomised. A crossover study must have at least 2 treatment groups. Because of its increased sensitivity and smaller variability, the crossover design requires fewer patients than does the purely parallel design to detect the same effect, however, it is not as robust as a purely parallel design, being adversely affected by patient withdrawals and missing data.
Cross-sectional	A study in which exposures and outcomes are measured at 1 point in time. These are nonexperimental studies with exposures/treatments measured rather than allocated. A cross-sectional study examines the relationship between disease/health-related characteristics and other variables of interest as they exist in a defined population at one particular time.
Double-blind	A study in which neither the subjects nor the investigators know the intervention to which the subjects have been assigned. A double-blind design is considered to provide the most reliable data from a clinical trial.

Epidemiological	This term encompasses cohort, cross-sectional, case-control and randomised-controlled study types.
In vitro	A study conducted on body tissue in an artificial environment (e.g. test tube). This term is only used by the Antibacterial service for antimicrobial activity studies.
Meta-analysis	This term is synonymous with systematic review. A systematic, quantitative analysis (usually of published literature) that statistically combines the data of multiple studies which address essentially similar research questions. Meta-analyses are typically used to increase the statistical power available to assess an intervention's effect on key outcomes and to better understand the size of that effect, as well as to assess the outcome effect on subcohorts. It is the quantitative statistical analysis of the results of the component studies that distinguishes a meta-analysis from a literature review or synthesis.
Model	A mathematical representation that illustrates the benefits/costs/outcomes of alternative decision choices, and offers the decision-maker a summary of alternative diagnostic and therapeutic options in a specific clinical condition.
Multicentre	A study that is conducted in patients from > 1 centre or institution.
Observational	A nonexperimental study in which the actual experiences of the groups being compared are simply observed.
Open	This term is synonymous with open-label and the opposite of double-blind i.e. subjects and investigators both know which intervention subjects are receiving. This term is only used for comparative studies that are stated to be open. For noncomparative studies that are stated to be open, prospective is used instead.
Parallel	Subjects receive 1 of 2 different treatments for the duration of the study. The parallel design is robust - i.e. tolerant to many kinds of problems that can occur in clinical trials, such as missed visits and missing data.
Postmarketing surveillance	A study conducted after the launch of a drug, to provide information on its use and on the occurrence of adverse events. Subjects are often general-practice patients. This type of study may be required by a country's regulatory authority in order to assuage concerns about tolerability. Alternatively, the study may be conducted voluntarily by the sponsor to obtain data useful to gaining greater market acceptance of the drug.
Prospective	A study conducted going forward from a point in time. This term is only used alone or in combination with multicentre. If the article states only that a noncomparative study is open, prospective is used instead.
Randomised	Subjects are allocated to treatment groups according to chance (e.g. using random number generation, sealed envelopes or a probabilistic rule). The purpose of randomisation is to produce comparison groups that differ only by the treatment allocated. Randomisation reduces bias in allocating treatment and ensures the validity of statistical significance tests.
Retrospective	A study using data previously recorded that was obtained either in a

	normal healthcare setting or, less frequently, in a controlled setting.
Sequential	A study in which subjects receive 2 or more treatments in the same chronological order e.g drug a then drug b. Results from the different treatments are compared.
Single-blind	A study in which either the subjects or the investigators are unaware of their treatment assignment, not both.
Survey	A nonexperimental study that systematically gathers information from a particular group. A survey may be conducted using a questionnaire, an interview or another method.

Adis Trial Design Score

Adis Clinical Trials Insight helps you make the most of the time you spend on scientific literature. Adis has developed a highly structured method to evaluate the adequacy of data provided by a therapeutic trial or economic analysis, based on the assignment of a numerical score to various elements of the study design, its conduct and adequacy of the report itself. Studies which evaluate quality of life include additional elements of assessment.

Adis Score for Therapeutic Trials

Elements evaluated to generate an Adis Score for Therapeutic Trials include:

Definition and pre-trial confirmation of disease severity in accordance with accepted classification criteria.
Adequate size of the patient groups to show a statistically significant difference between the therapeutic response of the drugs under investigation.
Comparability of patient groups according to accepted criteria and to the requirements of the disease being treated.
Therapeutic comparability of drug dosages used.
Duration of therapy in relation to the stated purpose of the trial.
Controls to reduce variation and bias.
Checks to confirm patient compliance.
Definition of therapeutic response in relationship to the type of patients and severity of their condition.
Relevance of the assessment of therapeutic efficacy to the stated purpose of the trial.
Adequacy and nature of side effect monitoring systems.
Relevance of the authors' conclusions to the results obtained.
The clinical relevance of the therapeutic response achieved.
Appropriate measures for quality-of-life assessment.

Adis Score of Trial Design for Economic Analysis

Elements evaluated to generate an Adis Score for Economic Analysis include:

Clear statement of aims of the study.
Clear definition of the study perspective, i.e. have the investigators taken the perspective of society, the healthcare system, insurers, hospitals, patients, or a combination of these?.
Validity of the treatment alternatives considered. Have all the relevant alternatives been considered?
Relevance of the outcomes measured to the stated purpose of the study.
Relevance of the costs measured to the study perspective.
Clear statement of the sources of cost and outcome data.
Appropriate discounting used.
Appropriate sensitivity analyses performed.
Full and accurate reporting of results, with appropriate discussion.
Relevance of the authors' conclusions to the results obtained.
The clinical relevance of results.

Adis Score Guide
86-100 Excellent trial, highly acceptable
71-85 Good to very good trial, most important elements adequate
50-70 Fair trial, some important elements inadequate
<50 Not acceptable, or results require confirmation by a better designed study

Occasionally, apparently well designed studies have a crucial flaw in design which substantially reduces the value of the results obtained. Such studies, according to the Adis scoring system, may still obtain a high score (>70). In this situation, despite the high score, the results and conclusions should be viewed with caution. An explanation of this is usually found in the "Adis comments" field.

Despite a low score, a trial may provide clinically relevant and important information on the practical experience with a drug. For example, some large scale, single drug controlled trials in general practice patients will not achieve a high score. An explanation of this is usually found in the "Adis comments" field.

Taken from: http://www.adisinsight.com/aClientServiceinfo/CTI%20Appendix.pdf