Appropriate Methods to Assess the Effectiveness and Efficacy of Treatments or Interventions to Control Cancer Pain

CLAUDIA BAUSEWEIN, M.D., M.Sc. and IRENE J. HIGGINSON, B.Med.Sci., B.M.B.S., Ph.D.

ABSTRACT

Pain is common in cancer patients. To ensure optimal pain management efficacy and effectiveness of new drugs and treatments have to be investigated in clinical trials. Efficacy trials such as randomised controlled trials (RCT) are experimental studies and estimate the maximum potential benefit to be derived from an intervention in ideal circumstances and under a controlled environment. RCTs are the only trial design to establish causal effects. A crossover study is a special type of RCT where patients serve as own controls. In efficacy studies the intervention and the control group should be as homogeneous as possible, confounding variables are controlled, bias is reduced, internal validity is high whereas external validity is low. Studies looking at effectiveness assess clinical practice and reflect real life circumstances. They rely high on external validity at the expense of careful controls, the study population is heterogeneous, confounding variables are examined. Cohort studies follow a group or groups of individuals with a common characteristic over a period of time to measure outcomes. Case-control studies start with the outcome and compare the characteristics of two groups of interest, those with the outcome and those without to identify factors which occur more or less often in the poor outcome group. Definition of outcome criteria is crucial both for efficacy and effectiveness studies and is often a primary problem. All clinical studies must use valid and reliable outcome measures.

INTRODUCTION

Pain is a common problem in patients with advanced cancer. As a multiplicity of new drugs and treatments become available an evidence-based approach can help to ensure optimal pain management. Efficacy and effectiveness have to be investigated in clinical trials. But these terms are rarely well defined and sometimes are used interchangeably or are not clearly distinguished. Yet they investigate different aspects of whether treatment will work. This review assesses the need to collect evidence of efficacy and effectiveness about therapies to alleviate pain, the appropriate study designs and their strengths and weaknesses. Efficacy is measured in randomized controlled trials or crossover trials whereas the measurement of effectiveness needs cohort or case-control studies. Issues such as study populations, analysis of data, retrospective and prospective designs, and outcome and outcome measures will be examined.

Efficacy Studies

Efficacy trials are experimental studies and estimate the maximum potential benefit to be derived from an intervention (e.g., treatment) in ideal circumstances and under a controlled environment (Table 1). They help to answer the ques-
tion whether a treatment works. In pain management they can serve as models which help to understand more about analgesic pharmacology. These carefully controlled trials maximise internal validity at the expense of generalisability.1 Efficacy in analgesic interventions is judged by the change of pain a patient reports.2

Randomized controlled trials

For testing the efficacy of a therapeutic intervention (e.g., an analgesic) randomized controlled trials (RCT) are an essential tool and the only trial design to establish causal effects.3 They are still seen as the gold standard in clinical research.4 A group of patients receiving a strictly regulated intervention under investigation (e.g., analgesics or radiotherapy for bone pain) is compared to a control group. The control group receives either standard treatment or a placebo. For example Ahmedzai and Brooks5 compared transdermal fentanyl with oral controlled-released morphine.

Through random assignment to study groups all potential “nuisance variables” of patients with cancer pain (age, gender, type of cancer, other symptoms, medical problems) are effectively equalized and selection bias minimized.6 Randomization reduces the risk that the differences between the study groups are the result of some systematic bias regarding how subjects are selected.6 However, this risk cannot completely be eliminated. There are different methods of randomisation (e.g., random number tables or use of computer-generated programs). If study groups differ in important variables stratified random sampling with random assignment within each level of stratification can be a solution to avoid severe confounding (e.g., stratification in neuropathic versus nociceptive pain). This procedure ensures equal distribution of crucial study variables.6 Nonrandomized trials suffer from strong selection bias (Table 1). They are much more likely to lead to enthusiastic treatment recommendations as compared with properly controlled trials. It is difficult to obtain a reliable assessment of treatment efficacy.3

The intervention, including the treatment, dose, route of administration, as well as the contrast between intervention and control studied, must be clearly defined in an efficacy study. Ideally the intervention, especially in a drug trial, is placebo-controlled. There is controversy about the use of placebo-controlled trials in palliative care.7,8 Ethical objections are raised as to whether it can be justified to use a placebo in pain syndromes that generally respond to treatment.9 Attrition rates may be high if this study design is chosen. To avoid this both the placebo and the active treatment group should have access to standard analgesic rescue doses.9 The overall analgesic consumption should then be a primary
outcome measure. Placebo “interventions” may not be appropriate when nonpharmacologic or invasive treatments are subject to investigation; an example might be coeliac axis plexus blocks. Many modern trials include a standard analgesic as positive controls that serve as a yardstick against which the magnitude of the analgesia produced by drug “X” is compared. Principally contamination with the control is possible and has to be avoided. If a crossover design is used a sufficient washout period should be planned.

Because RCTs are expensive, funding is sometimes difficult. Drug companies often sponsor RCTs that can potentially result in problems with publication bias (e.g., only trials that show potential benefits, for example “positive results” of the analgesic under investigation may be published).

Crossover trials

A crossover trial is one type of an RCT. Each patient receives at least two different treatments and serves as his own control. Ahmedzai and Brooks used a crossover design to investigate the use of sustained-release oral morphine and transdermal fentanyl and changed the treatment in each group after 15 days. The crossover design reduces experimental error that can occur because of random differences between subjects. One advantage of crossover trials is that substantially smaller sample sizes are required to show significant differences compared to standard RCTs. The most important potential limitation is the interference of effects of the drug given in the first phase with the second phase. This carryover effect happens especially in drugs with long half-lives (e.g., antidepressants, slow-release formulas). Planning a washout period between the two phases can help to avoid this effect but has ethical implications if the patients are to be left without treatment. A further problem is that in crossover trials a chronic pain condition or disease may change over time affecting the patient independent of treatment.

Study populations

In reality, patients with cancer pain are a heterogeneous group because they suffer from different malignancies at different stages. Patients’ experience of pain is determined by three main factors:

- Pain-related factors; pain characteristics and mechanism, prior analgesics.
- Patient-related factors; psychological and social factors.
- Disease-related factors; type of tumor, extent of disease, other symptoms, concomitant disease.

In efficacy studies the intervention and the control group should be as homogeneous as possible. To achieve homogeneity in this group of patients it is necessary to establish stricter inclusion and exclusion criteria than simply by type of primary cancer. Clinical and health parameters, symptoms and functional status reflect prognostic factors and can help to enhance homogeneity, and are part of the inclusion and exclusion criteria. However, multiple rigorous entry criteria may slow patient accrual and introduce selection bias. High rates of attrition might also be caused by disease progression or death as in the study by Ruttens et al. looking at the analgesic efficacy of radiotherapy for cancer pain, in which 24 of 75 patients included could not complete treatment. Because patients may drop out of the study after randomisation or fail to complete treatment this can affect the outcome. To reduce bias, all patients have to be analyzed in the group in which they have been randomized (“intention to treat” analysis).

Results of treatment in a particular study can only be generalized to the population of individuals from whom the sample was selected but not to a wider population. As a result, external validity is low in efficacy studies (Table 1). For example, if patients taking part in a study about tricyclics in neuropathic pain have to be under 65 years it is difficult to draw conclusions about side effects for an older population. Nonparticipants should be registered and evaluated to judge generalizability as defined in Table 1.

As in all research palliative care patients have to give informed consent to participate in the study. There are multiple ethical and practical issues concerning consent in palliative care, which is discussed elsewhere.

EFFECTIVENESS STUDIES

Once the efficacy of a treatment is established, the treatment needs to be tested in the real world of clinical practice. Studies looking at effectiveness assess clinical practice and reflect real life cir-
cumstances. They rely high on external validity at the expense of careful controls (Table 1). Effectiveness in palliative care is judged in terms of:

- Quality of life of patient and family:
- The adequacy of symptom control:
- The satisfaction of the patient and the family with the care received.14

From a public health perspective, what really matters is effectiveness.15 For studying effectiveness cohort (also called longitudinal) studies (Table 2) are more appropriate because they are based on observation and comparison of groups of people within free-living populations.

**Cohort study**

In a cohort study a group or groups of individuals with a common characteristic is followed over a period of time to measure outcomes.16 The studies can also be called longitudinal or observational. Common characteristics can be different exposures or interventions, levels of exposure, or presence of different factors. The period of time subjects are followed up varies widely depending on the time the outcome occurs after exposure. Cohort studies can be descriptive to measure frequency (incidence rates and prevalence) or analytical to measure associations.

Ideally a comparison or second group is included, but it is possible to follow only a single group. Unexposed groups serve either as internal or external controls. After formulating a clear study hypothesis the study group is selected according to the exposure or unusual experience they are undergoing. For example, Schug et al.17 followed 1070 cancer patients with pain over a period of 6 years and observed the use of morphine during that time. There is no randomization in a cohort study. This makes cohort studies much more vulnerable to selection bias and confounders. In studies of patients with cancer, attrition rates as a result of advancing disease or death may be high and compromise internal validity. Study subjects can be matched with a control subject for different variables.18 Alternatively confounders can be adjusted by statistical analyses such as stratification of the data or multivariate analyses.18

Interpreting cohort studies one also has to be aware of the potential serious implications of regression to the mean, which means that the observed effect could be caused by chance and not by change over time or the intervention.19,20

Cohort studies can be time consuming and expensive although they are often not as expensive as randomized trials. Sometimes costs can be lowered by collecting data in clinical practice.
Analysis of data

Completeness of follow-up data is vital. Schug et al. used total treatment days of morphine, types and dose of drugs taken, side effects, place of therapy, patient mood, and pain rating. Analysis of data includes the derivation of rates of specified outcomes among cohorts under study. Various groups of patients of the same cohort, different time points in care of the same cohort, or differences between exposed and unexposed cohorts can be compared. In double or multiple group cohort studies associations of exposure effects can be measured such as relative risk, attributable risk, and rate ratios (Table 3). Descriptive cohort studies serve to measure incidence rates.

Retrospective versus prospective design

The classic cohort study design is prospective but retrospective designs (historical cohorts) are also possible. In a prospective design, the methods of data collection and cohort are identified in advance and followed over time following a research protocol. In a retrospective (or historical) cohort study, a cohort is identified but the analyses use data collected as part of routine care, or extracts longitudinal information from clinical or other records. Thus the data collection in a retrospective cohort study can be subject to many biases, although it is often less expensive than a prospective study. Even if in a retrospective study the outcome is not known, the presence or absence of the exposure of interest is important. Prospective designs allow multiple measurements over time and more accurate measurement of exposure and potential confounding variables.

Case-control studies

Another form of epidemiologic study is the case-control study. These start with the outcome and compare the characteristics of two groups of interest, those with the outcome and those without (e.g., pain controlled and not controlled) to identify factors which occur more or less often in the poor outcome group. Cases can be matched to reduce the effect of potential confounding variables. However, matching does not on its own eliminate confounding. Statistical adjustment with stratification or multivariate analyses is often still necessary. Case-control studies are retrospective. They are relatively inexpensive and can provide quick results. However, their main use is in the investigation of public health infection outbreaks and they are rarely used as yet in palliative care.

Factors influencing effectiveness

Effectiveness depends on:

- Efficacy;
- Diagnostic accuracy in the field;
- Health professionals' compliance;
- Patients' compliance; and
- Coverage.

Failings in any of these factors will consequently reduce effectiveness. To ensure diagnostic accuracy practising clinicians must correctly identify those who may benefit from treatment. Errors may occur if patients who may benefit from the treatment are not diagnosed (e.g. if patients with severe pain are not receiving morphine) or in contrast, if indications for treatment are broadened with the result that patients who are less likely to benefit receive the treatment. Further issues that can influence effectiveness are

- Fears about prescribing opioids for severe cancer pain both by physicians and patients;

<table>
<thead>
<tr>
<th>Table 3. Analysis Possible from Observational Studies</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Incidence rate</strong></td>
</tr>
<tr>
<td><strong>Odds ratio (cohort study)</strong></td>
</tr>
<tr>
<td><strong>Odds ratio (case-control study)</strong></td>
</tr>
<tr>
<td><strong>Relative risk</strong></td>
</tr>
<tr>
<td><strong>Attributable risk</strong></td>
</tr>
<tr>
<td><strong>Rate ratios</strong></td>
</tr>
</tbody>
</table>
• Dislike of drugs because of taste by many patients;
• Side effects of opioids, such as constipation or dry mouth;
• Coverage and cost-effectiveness of a drug; and
• Costs, especially when patients have to purchase drugs themselves.

Emphasis should be put on monitoring effectiveness through development of regular audit and national and international guidelines such as the European Association of Palliative Care (EAPC) recommendations on morphine and alternative opioids in cancer pain.24

OUTCOME AND OUTCOME MEASURES

Definition of outcome criteria is crucial both for efficacy and effectiveness studies and is often a primary problem.25 Multiple outcomes can be measured routinely or specially recorded. Outcomes in analgesic trials can include

• Pain intensity and pain relief;
• Quality of sleep;
• Acceptability of therapy;
• Dose of rescue medication; and
• Side effects.

Researchers have to be aware that in cancer and palliative care patient outcome may be compromised by other factors such as disease-related factors, personal circumstances or concurrent use of care.10 Furthermore the treatment success often reflects the agenda and relative values of the different parties evaluating treatment.25 Patients might define reduction of pain severity as success whereas the researcher looks at improvement in functional status. Therefore, a combination of outcomes should be used including pain and functioning preferably with a mix of subjective and objective measures.6 Side effects should also be assessed.

Because pain is a subjective, complex, and multidimensional experience there are no objective methods to measure pain.26 However different subjective descriptions can be used. Pain intensity is recognized as one of the most relevant dimensions which can be measured in a reliable and valid way.26 All clinical studies must use valid and reliable outcome measures that are validated by prior research. Different pain measurement tools exist (Table 4). The McGill Pain Questionnaire27 and the short form of the Brief Pain Inventory28 are recommended for assessing pain in research by the Expert Working Group of the EAPC.26 Interday and intraday variability of pain must be taken into account.6 Researchers must be aware about the difference between patient and physician ratings as therapists rate patients often as much more improved than do patients themselves.25 Because there is no one correct outcome measure the question remains open how to determine the efficacy of various treatments when diverse measures and criteria are used and frequently combined to demonstrate treatment success.25 Discussion arose around the question whether a 50% reduction in baseline pain scores is the definition of a successful pain intervention. Farrar et al.29 showed that a reduction of approximately two points or a reduction of approximately 30% in a numeric rating scale (NRS) represented a clinically important difference.

### Table 4. Pain Measurement Tools

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Scale</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unidimensional</td>
<td>Visual analogue scales (VAS)</td>
<td>“Ideal” scale, independent from language</td>
</tr>
<tr>
<td></td>
<td>Categorical verbal rating scales (VRS)</td>
<td>Variety of choices from 4-grade to 15-grade scales</td>
</tr>
<tr>
<td></td>
<td>Categorical numerical rating scales (NRS)</td>
<td>Easier to apply, better compliance</td>
</tr>
<tr>
<td></td>
<td>McGill Pain Questionnaire</td>
<td>Global scores, subscale scores reflecting sensory, affective and evaluative dimensions of pain</td>
</tr>
<tr>
<td>Multidimensional</td>
<td>Brief Pain Inventory</td>
<td>Simple, easy to administer, information about history, intensity, location and quality of pain</td>
</tr>
<tr>
<td></td>
<td>Memorial Pain Assessment Card</td>
<td>Only available in English</td>
</tr>
</tbody>
</table>
CONCLUSION

In medicine, RCTs are still seen as the gold standard and other research designs are commonly less accepted. But finding the right research design and strategy for a study is not so much a question of the method but of the research question being asked. Assessment of interventions to control cancer pain varies according to the clinical situation. If the research question is whether a new analgesic is different from standard treatment, then an experimental study such as an RCT or a crossover trial is the best trial design. Attention must be paid to patients who drop out after randomization. These patients need to be included in an “intention to treat” analysis.

After showing efficacy of a treatment or intervention the effectiveness has to be tested in real-world circumstances in observational studies. Cohort studies follow patients over time and measure outcome, and are critical. Observed changes have to be interpreted with caution because of regression to the mean and bias.

Experimental studies have high internal validity but observational studies are needed to ensure external validity, and are more clinically representative. In both, outcomes appropriate to the research question have to be defined and appropriate measures chosen. Differences between efficacy and effectiveness should be explored, as these can be caused by problems regarding individual patient factors, professionals’ knowledge and attitude, financing of drugs, or public health concerns.

ACKNOWLEDGMENT

This paper was based on an assignment within the Master’s of Science course for Palliative Care at King’s College London.

REFERENCES


Address reprint requests to:
Claudia Bausewein M.D., M.Sc.
Interdisciplinary Centre for Palliative Medicine
University Hospital of Munich–Grosshadern
Marchioninistr. 15
81377 München
Germany

E-mail: Claudia.Bausewein@med.uni-muenchen.de