In: Palliative Radiation Therapy
Editor: Alysa M. Fairchild
ISBN: 978-1-63463-634-6

Chapter 6

Methodological Issues and Challenges in Performing Clinical Trials in Palliative Radiotherapy

Jackson S. Y. Wu MD MSc FRCPC
Department of Oncology, Tom Baker Cancer Centre, University of Calgary, Calgary, Alberta, Canada

"It is certainly unethical to conduct an invalid experiment." ~ Sir Bradford Hill
"Science is our business to think, not to codify thoughtlessness." ~ Lemuel A. Moye
"…research is defined as an undertaking intended to extend knowledge through a disciplined inquiry or systematic investigation." ~ Tri-Council Policy Statement 2010

Abstract

In the design and execution of clinical trials for the palliative cancer population, there are challenges that are common to interventional trials in palliative care, but there are also inherent limits of focal external beam radiotherapy (RT) as an intervention that demand thorough consideration and deliberations. In the palliative cancer practice setting, two patient-related factors often demand added complexity or practical compromises in trial design and conduct: heterogeneity and attrition. Other design-related issues include strategies to define and obtain clinically meaningful endpoints efficiently, choice of patient reported outcomes and validated instruments, and knowledge of minimal clinically important difference. In the planning of sample size and power requirements for statistical inferences, the type of primary endpoint, be it singular or composite, requires substantial justification that should address therapeutic objective and clinical value, while matching the expected or putative mechanism of effect associated with the intervention. Given a myriad of subscales with common quality of life instruments, pre-specified complementary palliative effects as secondary endpoints will inform potential real-world value, as long as the study's primary objective is met. In addition to challenges of conducting trials in the palliative care setting, palliative RT trial designs are challenged by ambiguity in primary or intermediary therapeutic objectives. For example, symptom relief may be achieved with or without cytoreduction, and local control may be

demonstrated with or without impact on disease progression. Because focal RT can be delivered to different target lesions within a given patient simultaneously or over time in several courses, the mechanism of therapeutic advantage should be clearly reasoned and appropriate statistical methods for correlated outcomes data articulated in advance. Logistical issues such as ability to provide eligibility screening, consent, study registration, treatment planning and delivery in limit time or visits can lead to ineligibility or dropout, particularly for those patients with high symptom burden. Finally, the value of delivering ablative doses of RT in the palliative setting might be demonstrated through carefully designed and executed randomized controlled trials, particularly with respect to an appropriately defined control arm, clinically meaningful follow up time, attention to adverse events and inclusion of resource utilization data to estimate incremental costs. While there are many challenges to conducting palliative RT trials in the palliative cancer population, they are not insurmountable. This chapter will review several basic issues in statistical reasoning fundamental to research methodology, and discuss those challenges that the investigator should consider when preparing a study proposal.

Keywords: Radiotherapy; clinical trial; methodology

Introduction

Innovations in science and technology have contributed much towards modern medicine and society as a whole. But the practice of clinical medicine relies on more than technology; the ethical physician must take into account many sources of knowledge for clinical decision-making, ranging from basic anatomy and pathophysiology to his or her own experience through managing the sick patient, plus an overwhelming abundance of abstracts presented at scientific meetings and published articles, let alone opinions from well-meaning colleagues, patients and families. In this complex information environment, a necessary if not essential safeguard against misguided decisions or ill-conceived judgment calls is the process of learning through the methodical design and execution of a disciplined research study. Clearly-formulated hypotheses and defined objectives, appropriate measures of clinical effects - both benefits and harms - and a dedication to scientific principles are key ingredients that allow valid inferences to be made from a sample to the target population. While there are many textbooks that provide essential background training in research methods for the keen clinician researcher, this chapter will offer an additional source of study for those eager to explore and test new strategies to palliate cancer patients through a well-designed, grant-worthy clinical research proposal.

The issues reviewed herein are by no means exhaustive; rather they highlight those aspects of study design and execution that tend to plague research efforts, challenging both the internal and external validity of the results. Many, if not all, of these issues need to be addressed during the design phase of the study, and they are common to all clinical trial efforts in general. Additional considerations concerning performing clinical trials in the palliative care population are discussed, and some perspective specifically on palliative radiotherapy (RT) is offered.

Many of the methodological issues discussed are also relevant to non-experimental study designs (e.g. retrospective case series or cohort studies, administrative data studies). Readers planning a chart review-type study may also find clarifications or suggestions on pertinent

issues in this chapter. Familiarity with these methodological issues will also facilitate critical appraisal of a study proposal and a clinical trial report in manuscript or published form. For better appreciation of clinical trial methodology, among other interesting aspects of biostatistics, the reader is strongly encouraged to consult the highly readable book *Statistical Reasoning in Medicine* by Moye [1] and references cited throughout this chapter. Of note, this chapter concerns issues related to quantitative methods, which is not to suggest that other methods, namely qualitative or mixed methods, are not important and relevant approaches to address scientific inquiries, but are simply outside of its scope.

In the Beginning…There Is a Hypothesis

Ideas need words and language to be conveyed, debated and accepted. Choosing an appropriate study design begins with a clear statement of hypothesis and measurable study objectives. (One might add that before a hypothesis is the theory.) The choice of words in these statements implies intent, and, more importantly, the framing of the hypothesis and objectives. Not infrequently among study proposals and abstracts in scientific meetings, the words "to determine" or "to identify the causes of" are used. Clinician researchers should be mindful that while most research efforts seek to demonstrate or discover causal relationships, for example, between a putative diagnostic or therapeutic intervention and the outcome of interest, or between an exposure and progression of a disease, the objective statements "to determine" or "to identify" imply a study aiming at providing strong evidence for causal relationships. This is a tall order and a major undertaking that calls for meticulous experimental efforts and/or observational evidence, most often in the form of a randomized controlled trial (RCT) or, respecting inherent limitations, population-based studies that accounts for known and potential confounders. By practical and design limitations, the purpose of single-centre observational case series or multicentre experimental phase 1 or 2 trials will not meet such a goal. Of course, it would still be semantically correct to say one could "determine" patterns and associations solely within his particular study participants without inferring the truthfulness of those patterns outside of his study. But most readers would be interested in analyses and interpretations that *could* be applicable (generalizable) to their setting, and it is best to avoid high level study objectives unless one is so prepared to undertake. Table 1 provides a brief review of conventional clinical trial designs and the framing of their respective study objectives.

When it comes to hypothesis and objectives, the enthusiastic investigator begins with the desire for his study "to demonstrate the benefits of an innovative treatment X", often with the view that one-sided hypothesis testing would suffice in the study design. Indeed, it would not be difficult to design a study to "demonstrate benefits", for example by ignoring (or wrongfully "censoring") subjects who succumb early or lose to follow up, or by comparing the experimental intervention to suboptimal control conditions. One problem of holding a strong "one-sided" view of an intervention *under question* is the risk of introducing subtle biases in the design of the study. The investigator should make every effort to understand the reasoning behind two-sided hypothesis testing (Moye devoted an entire chapter in his book on this one topic [1]), recognize the potential of both putative benefits and unanticipated harms, and seek input from clinicians and researchers of other disciplines who would have more

neutral observations and understanding of the study implications. Such a process would enhance the objectivity of study design and the value of a grant proposal.

Table 1. Type of study design and scope. *Avoid using the action term "to correlate" an intervention or exposure with an outcome of interest. Abbreviation: RT – radiotherapy

	Objective	Scope	Sample size Justification	Analogous Observational Study Design
Phase 0	To provide pre-clinical data, to enhance understanding of mechanism of action, potential toxicity, theoretical therapeutic advantage	Exploratory, pre-clinical (no human participants); basic science and translational discovery research; toxicology studies; in-vitro, in-vivo and/or animal models; RT dosimetry studies (phantom or computer models)	Depends on the nature of study objective	n/a
Phase 1	To evaluate maximal tolerated dose and dose limiting toxicities (organs); to assess pharmacokinetics properties	Exploratory, risk assessment in individuals (previously unknown or uncertain because of dose-schedule modifications)	Dose-escalation methods [ref.33]	Clinical anecdotes, case reports, tumor board discussions; small clinical practice case series
Phase 2	To estimate efficacy* (treatment response), to further evaluate toxicity profile and severity; to assess adherence to protocol logistics so as to inform and refine future phase 3 study design	Exploratory preparation, estimation of event rates (treatment response and adverse events) to motivate phase 3 trial and inform sample size, or, alternatively, to reconsider line of query (if efficacy estimate does not meet expectation;	Simon two-stage design; typically based on expected response rate and precision, e.g. expect 60% response +/- 10% [refs.34,35]	Larger clinical practice case series; institutional case series or cohort studies
Phase 3	To determine cause-effect relationship; to estimate efficacy* (effect size) of intervention in the population; to assess issues of risk-benefit	Confirmatory trial of primary study objective; supporting evidence from secondary objectives; judgment (interpretation) on real world value may include cost-effectiveness, quality of life and other pre-specified secondary analyses	Effect size deliberations among stakeholders	Population-based cohorts, parallel cohorts, case-control designs controlling for confounders (but not direct measures of cause-effect relationships)
Phase 4	Post-marketing observations (usually voluntary adverse events reporting)	Further information on adverse effects not observable in phase 3 setting (based on sample, not population)	Convenience	Quality assurance studies; practice surveys
Pilot / feasibility study [36]	To identify and address issues that could occur in the subsequent confirmatory trial	To evaluate feasibility of protocol implementation before full-scale phase 3 trial is launched; to investigate potential mechanisms of efficacy	Judgment, aim-specific considerations (see Moore)	Practice surveys

Sometimes clinicians have difficulty articulating a study hypothesis within a single-arm trial design where there is no contemporaneous comparator. It is important to appreciate that a hypothesis could be stated with respect to a hypothesized point estimate and its statistical precision (not limited to comparative hypothesis testing between or among groups), where the expected outcome, based on *prior experience* or reasoned expectation of the new intervention, would produce X% success rate with plus or minus D% confidence limits (e.g. response or survival rate with 95% confidence interval at pre-specified follow up assessment time).[2] In a single-arm trial, the choice of the success rate and the acceptable variability (both are needed in order to compute the sample size) must be defined and justified by a level of clinical relevance deemed sufficient to motivate a subsequent phase 3 study or, in special circumstances, a change of clinical practice. If the objective (i.e. expected efficacy) is not met after careful study execution, the investigator should also be prepared to discuss how the concept or theory that led to the hypothesis might have be reconsidered.

In clinical trials where the comparator is an active control, a decision needs to be made regarding one-sided over two-sided hypothesis testing when both interventions are reasonably understood and have similar risk profiles. One-sided hypothesis testing in such a situation can be done within a non-inferiority design framework, where the intervention under question - likely with improved ease of administration, better toxicity profile or tolerance, or lower cost - is compared to another accepted, active treatment, and the expectation of chance experimental outcome (i.e. P-value) is not symmetrical. In a non-inferiority trial, the hypothesis is that treatment under question is not worse than standard by more than a pre-specified margin, irrespective of its potential superiority.[3] Because hypothesis testing is one-sided, the sample size needed is smaller, but often not as much of a reduction as what one might expect (going from two-sided to one-sided hypothesis does not reduce the sample size by 50%, for example).[1] It should be noted that a non-inferiority trial demands a level of rigor no less and perhaps more so than the typical two-sided comparative trial (i.e. “superiority trial”), because of nuances that tend to bias the results towards non-inferiority. Table 2 highlights some basic features of non-inferiority and comparative trial designs. The researcher must be alert to the nature of the intervention under question when choosing one-sided or two-sided hypothesis testing. If the intervention under question is more intense compared to the standard intervention (e.g. dose-intense or new combination therapy), choosing a non-inferiority design is not rational; clearly one would not compare the performance of a newly designed SUV to the old family station wagon just to say the new SUV is “non-inferior”.

Central to some of the confusion among clinician researchers in developing a study proposal or clinical trial protocol are some common misconceptions around the nature of statistical inference and hypothesis testing. Before we dive into the discussion of clinical trial issues in the palliative oncology population, we must take a step back to get them straight.

The P-Value Fallacy and Errors of Random Research

Among clinician-driven research studies, scientific meeting abstracts and publications, there is a preponderance, if not obsession, towards P-value based interpretation of study

results. A typical research study presented at a national society meeting might say that patient and treatment factors a, b, c…j were collected, and factors b,c, e, g, h, j contained sufficient data to be tested for their association with outcomes L, M, N using statistical procedures P, Q, and R, in a convenience sample of size S, reduced from a referred population of already self-selected cases because of loss to follow up. After additional multivariable procedures, in which some of the baseline factors were included based on their P-value significance in univariate analyses, factors g and j were "significant" predictors of the outcome of interest. The abstract goes on to conclude that factors g and j should be incorporated into new guidelines for baseline patient evaluation of the condition in question.

Table 2. Basic features of comparative and non-inferiority trial designs. Abbreviations: Δ - pre-specified measure of meaningful effect (e.g. 10% difference in response rate at 2 months); M - non-inferiority margin

Trial Design	Hypothesis	Nuances
Comparative/ Superiority	H0: Control – Experiment $\leq \pm\Delta$ Ha: Control – Experiment $> \pm\Delta$	Placebo-controlled trials; intention-to-treat analysis is conservative: imprecisions in trial conduct (deviations, measurement error) tend to reduce effect size (dilute results) and bias towards H0.
Non-inferiority	H0: Control – Experiment $\geq M$ Ha: Control – Experiment $< M$	Active control trials (e.g. "me too" drugs); imprecisions in trial conduct (e.g. missing data) tend to reduce effect size and bias towards Ha or inconclusive results; *M=non-inferiority margin*: a matter of clinical judgment and consensus at trial set up. Sample size is justified for precision (i.e. confidence limit < *M*), which increases sample size, though offset by one-sided hypothesis testing.
Equivalence	H0: Control – Experiment $\geq \pm\Delta$ Ha: Control – Experiment $< \pm\Delta$	Active control, rarely done (mostly non-inferiority question). Imprecisions problems similar to non-inferiority design. Sample size is justified for precision (confidence limits $< \pm\Delta$), which increases sample size.

While biostatisticians and research methodologists would be gravely concerned about the nature of the missing cases, the lack of an *a priori* hypothesis and a pre-specified multivariable model, and the absence of sample size considerations, among other things, what is shocking is the common acceptance among clinicians that the P-value result, when "significant", confers all manners of legitimacy to the study design and execution. Not only does the P-value accurately account for the sampling error or variability (as it is meant to do), it is also expected (erroneously) to resolve all potential issues of bias and confounding, attest to appropriateness of sample size, identify a cause-effect relationship, signify the effect size ("significant difference") as relevant and meaningful, and validate the study results and

interpretations. It is as though once "statistically significance" is declared, something truly "significant" has been found, and we bend over backwards to acknowledge and accept its presence, instead of thinking over its validity and relevance. When one reads through many such scientific meeting abstracts, and, not infrequently, peer-reviewed publications, it is sometimes hard to draw a line between folly (wishful thinking) and fallacy (error in reasoning) in the analyses and interpretations offered by some clinical researchers.

The two main concepts that need to be clarified to many clinicians and to those with limited research methodology training are related to the purpose and meaning of statistical inference and the quintessential features of a structured and disciplined research study. Statistical inference can be defined as "a conclusion that patterns in the data are present in some broader context…justified by a probability model linking the data to the broader context."[4] (Or simply put, extrapolating results from a sample to the population with a formal measure of the probability of error.) The two approaches to statistical inference are hypothesis testing and estimation, where estimation is the evaluation of effect size or average difference, namely the point estimate, along with the variability of the effect size, namely the standard error or interval estimate. Users of statistical analyses must recognize that results are *observed* in a sample and *inferred* to a larger population with the assumption that the group of subjects assembled for the study (supposedly random) is an appropriate representation of the target population and not skewed in some important way.

In statistical inference, whether the results are presented as P-values or as effect estimates within the structure of an appropriately executed research protocol, the test statistics serve as a measure of how likely the results could have occurred *by chance alone*, and not a result of the presupposed relationship between exposure (or intervention) and outcome, given conditions of the study design. The measures of variability (confidence limits, P-values) quantify formally the probability of those results being produced through random sampling or random allocation, but they do *not gauge the likelihood of the truthfulness* of those results in the population. These two issues, the probability of observing results by chance and the probability of the hypothesis being true or false, are separate constructs. Despite that unknown degree of truthfulness of what is hypothesized, decisions guided by proper statistical inference will, in the long run, produce the desired results in the population and will help people make the right choices more often than not. Misapplication of statistical inference, however, corrupts the long term view offered by statistical reasoning. Table 3 reviews the common products of statistical inference and their limitations.

While the P-value and effect estimates arithmetically capture those measures of variability from sample to sample, they provide neither information nor judgment on the enrollment (the kinds of patients referred or present to a particular clinician investigator) or eligibility (who is included/excluded per protocol) of the patients that could limit the generalizability of the observed results, nor an evaluation of potential deficiencies in study execution, for example, non-adherence to treatment procedure, loss to follow up and other reasons for missing data [see below]. Additionally, the clinical relevance or importance of observed results is not represented in the P-value itself. Investigators must critically identify and resolve those issues of study design and execution that might invalidate the study's generalizability apart from the P-value and statistical products. Goodman and Silva-Aycaguer have discussed many misuses or misconceptions of P-values, such as the incorrect presupposition of it being a measure of the truthfulness of the hypothesis, and the tendency for researchers or readers to reduce clinical information into significant vs. not significant as a

cognitive shortcut to evaluate evidence (zero or one thinking).[5-7] Moye provides the most reader-friendly discussions – albeit over three chapters – of the epistemological and historical origins as well as misapplications of this over-interpreted arithmetic product.[1]

Table 3. Products of statistical inference. * Note meaning of word usage: "effect" here denotes generic object of measurement with respect to observed results, and does not necessarily indicate "effectiveness" of an intervention with respect to clinical applicability or generalizability

Products	Measure	Requirements
Effect direction*	Estimate effect direction for or against the desired effect in the target population	Representative sample, pre-specified effects of interest (primary or secondary), adherence to protocol procedures & measurements, sample size justification.
Effect magnitude	Estimate effect size, effect (point) estimate, average difference in the target population	
Effect variability	Estimate sampling variability, standard error of effect estimate (interval estimate), range of potential effects (e.g. 95% confidence limits), given data	
P-value	Probability of acquiring by chance a result as extreme as or more extreme than what is observed, assuming null hypothesis to be true, based on the random sampling of subjects from a target population or random allocation in an experiment	Same as above. Caveat: Statistical significance can be a result of (1) bias, (2) correct hypothesis, or (3) chance. P-value provides no information on (1) or (2).

Apart from understanding the meaning and implications of statistical inference, the clinician researcher has to be aware of another fundamental requirement in order for statistical inference to be valid, whether the research is experimental or observational. The variability (sample error) computed by statistical methods is accurate only when the *examination* of that measure is *prospectively* declared (i.e. pre-specified hypothesis). This prospectively declared hypothesis includes pre-specifying the variables for a multivariable model. Not infrequently, however, investigators omit the construction of a thorough statistical analysis plan – consider it a fair trade contract between the investigator and the data – prior to initiating data analyses. As a result, multiple statistical analyses, sometimes in the dozens, are conducted *ad hoc*, with confirmatory findings declared falsely by mere "statistical significance" among those results. Sometimes those "significant" results are output of secondary or exploratory objectives, discussed as the primary research outcome *in retrospect*, when the primary study outcome is "non-significant".

In all of these circumstances, the investigator has carried out "random research",[1] in which the interpretation of "significance" or the sequence of conducting statistical procedures are motivated by the results of data analyses ("data-driven" analyses), not through the significance threshold specified by *a priori* reasoning. The problem with data-driven research

is that the quantification of *variability* through most statistical procedures does *not* account for the added levels of variability due to an assortment of plausible or convenient endpoints and hypotheses.[8] Similarly, anytime a researcher deviates from or alters the protocol's conduct and intended analyses (discordant research), new sources of variability are introduced, and the conditions necessary for valid statistical inference are seriously compromised or simply voided.

Some clinicians and researchers may expect significance threshold adjustments methods, such as the Bonferroni correction and the Hochberg sequential procedure, to allow multiple statistical tests or comparisons to be performed with legitimacy, as they supposedly account for "familywise error rate" and provide safeguards against type 1 error inflation.[9] Although these procedures are helpful when prospectively declared, they do not fully repair the damage to the "link between the endpoint and the significance threshold", as the type 1 error threshold that is fundamental to hypothesis testing is no longer fully controlled by the investigator, but shared with the data. Moye provides an approachable discussion of a structured approach, or "analysis triage", to pre-specifying type 1 error in situations requiring multiple testing with limited primary endpoints.[1] A more detailed examination of the subject is reviewed elsewhere.[10]

A more serious problem with data-driven analyses is with automated variable selection procedures (e.g. backwards stepwise variable selection), even though the researcher has the ability to set and change the in- and out- P-value thresholds (nonetheless arbitrary) within these procedures. Reliability problems associated with such procedures have been elaborated.[11-13] A similar data-driven analysis approach is with the use of bivariate associations (commonly "univariate analyses") to select those variables with "significant" P values to be included in multivariable analyses.[14][1] While these latter examples of flawed approaches to statistical inference are mainly seen in observational studies, they illustrate the common problem of data-driven analyses and interpretation, and are examples of incomplete study designs or "random research".

Attention to Detail in Structuring a Clinical Trial Protocol

Making statistical inference – deciding what might be true in the population based on a relatively limited sample – demands rigorous efforts in study design, planning and execution. An essential step is to finalize the statistical analysis plan before beginning the analyses. To respect the basic rules of statistical inference and to avoid the fallacies of data-driven, random or discordant research, it is important if not essential for investigators to gain a thorough understanding of the study hypothesis, deliberate with other experienced researchers on the scope of the research question, and to prioritize study objectives, such that appropriate conclusions can be drawn based on valid statistical inferences. This calls for clarity and a single well-defended primary objective and endpoint, or very limited endpoints (more about multiple and composite endpoints later), thorough considerations of appropriate patient selection criteria, and disciplined trial and data collection. It is critical that the study

[1] NB: Multivariable analysis is not to be equated with multivariate analysis; they are different analytical procedures.

conclusion is drawn – whether the study hypothesis is rejected or not rejected, or the study objective has been met or not met – based on the primary endpoint. But sometimes primary effect size estimate in a randomized trial can be better elucidated by adjusting for predictors of the outcome in a multivariable analysis, which must be also fully pre-specified (more discussion below).

Secondary endpoints, which serve as supporting evidence to the primary outcome/endpoint or to provide information about mechanism of action, should also be declared in the study design and their data rigorously collected, but the statistical inferences made *cannot* be deemed conclusive or practice changing, as "the play of chance could have considerable influence" on the "significant" finding, a chance not completely accounted for in the P-value.[15,16] Exploratory endpoints concerning ancillary data explored for various reasons depending on the type of clinical trial should be considered descriptive (without statistical inference). A structured approach to the examination of study outcomes, by "triaging" the type 1 error thresholds among limited pre-specified endpoints, is more interpretable than relying on statistical procedures to reveal "significant" relationships among a myriad of them.[1]

Depending on the nature of the scientific query, the development of a research proposal, be it a non-experimental design or a clinical trial protocol, can be arduous. The aim of the protocol is not so much to develop a perfect study, but to anticipate the challenges and acknowledge what areas are justifiable compromises against others, yet still support the overall study goal. Studies experience expected and unexpected obstacles (such as slow accrual or new technical specifications) or change in concept (especially new information from other studies). It is easy to presume that "protocol amendments" are a fact of life, but amendments (apart from administrative updates and clarifications) that alter the original study design should not alter the core hypothesis or how the hypothesis is framed. Great care and concern needs to be given both at the initial trial design stage (to avoid the need for amendments) and when amendments are being discussed. One should not assume that amendments will preserve the generalizability of the trial.

After the investigator is clear on what should be expected within a research endeavor, he or she is likely to encounter some methodological issues, which can be particularly troublesome in palliative care research in the cancer population. We will discuss some of these issues next.

Methodological Issues in Studying Palliative Cancer Populations

It is assumed that the intent of "palliation" is to moderate the intensity of the clinical ill-effects of a disease. It remains debatable whether a disease without current signs or symptoms requires "palliative" therapy, even as new treatments are showing life-extending efficacies for some patients with limited, though nonetheless "incurable", metastatic cancer. The methodological issues discussed in the following sections primarily address palliative interventions for the purpose of symptom control or relief.

Issues Related to Patient Heterogeneity

In designing a clinical trial protocol for cancer patients with advanced or relapsed disease that is not curable, a typical question concerns the degree to which a new intervention is superior or non-inferior to a control (i.e. usual care). In the palliative cancer population, a myriad of factors complicates the evaluation of symptom relief and ascertainment of cause-effect relationship between study intervention and outcomes. A fundamental aspect is defining the target population for which the intervention is hypothesized to provide an improved risk:benefit ratio. As this population can be extremely heterogeneous, from the asymptomatic, low disease burden, fully active patient to the frail patient with progressing cancer and acute-on-chronic symptom exacerbations, a palliative intervention has to define not only the appropriate symptom for which the intervention is indicated, but also the factors and potential contraindications that would render the intervention suboptimal, or potentially harmful.

The researcher would take great care to consider baseline factors that characterize the palliative population and how they might bias response estimates (in a given study arm) and affect the effect size (difference between arms) in the context of a clinical trial. Table 4 outlines some of these factors.

Table 4. Clinical factors that could bias efficacy and effect size estimates in palliative care trials. Abbreviation: n/a - not applicable

Factors	May Bias Estimation of Response and Assessment of Tolerance	May Bias Effect Size
Patient-Related		
Constitutional status	Aging, asthenia, weight loss and frailty are present to varying degrees in the sample depending on practice type or referral patterns; may reduce treatment effects compared to healthier patients, especially if reduced compliance; toxicity may be more frequent or pronounced due to diminished capacity for tissue recovery, interfering with response evaluation	If new treatment is prone to compliance problems (drop out due to toxicity, incomplete prescriptions), especially in participants with poor constitutional status, effect size can no longer be attributed to random allocation alone (confounding due to imbalance post-randomization)
Performance status	Efficacy may be more apparent in better performing patients as their evaluations are more reliable; efficacy estimate likely influenced by distribution of patients by performance levels	Same as above for study participants with lower performance status; also higher performance status patients receiving new treatment become more able to receive additional cancer therapies, magnifying apparent effect size for that group (false attribution of causality)

Table 4. (Continued)

Factors	May Bias Estimation of Response and Assessment of Tolerance	May Bias Effect Size
Comorbidities	Complicates capacity for tissue recovery and tolerance to new treatment; more complex pathophysiological influences; potentially reduce treatment efficacy; propensity to suffer unrelated adverse events, interfering with efficacy and toxicity assessments	Same as above for those with compromised constitutional status
Psychosocial context	Palliative cancer patients may become more dependent on family caregivers during a course of treatment; the lack of support may compromise treatment compliance and follow up assessments	Potentially moderates the severity of adverse events for those with strong social support (e.g. able to avoid hospital visits even if adverse effects are more severe in new treatment arm)
Disease-Related		
Cancer type	Some cancer types are more sensitive or responsive to therapy (e.g. palliative radiotherapy in breast cancer better response than poorly differentiated non-small cell lung cancer); palliative care trial involving all cancer patients will be more difficult to generalize if the selection of participants is not truly random; for cancers with many therapeutic options, sequential therapies will affect efficacy (and toxicity) estimates	For cancers with many therapeutic options, new treatment may allow participants to receive additional therapies, magnifying apparent effect size for that group (false attribution of causality)
Extent of disease	Diseases involving multiple organs limits potential efficacy because of multiple targets with additional variability of responsiveness to treatment; symptom severity complicated by disease burden, especially with bulky disease; propensity to develop acute complications and unexpected events	May reduce effect size
Relapsing versus de novo cancer	Natural history or responsiveness to treatment for relapsing cancers may differ from de novo advanced or metastatic cancer, especially if long disease free interval; efficacy measures may also differ between these two groups	n/a
Pathophysiology of symptom manifestation, symptom cluster	A given symptom may have multiple causes (due to cancer, frailty, late effects of cancer therapies, comorbidities, etc). If symptom-of-interest is not specific to the target pathophysiological mechanism, efficacy may be masked; symptom cluster may interfere with assessment of a single or index symptom (i.e. more difficult to see “relief” in a patient with complex symptom manifestation)	Problematic if significant co-interventions are involved, which would dilute the effects of study treatments (both arms), reducing effect size
Treatment-Related		
Prior cancer therapies	Progressing cancers after prior therapies may harbor resistant clonogens and their symptoms may be more resistant to therapeutic effects; may reduce apparent efficacy	n/a
Co-interventions for the same versus other symptoms	Increases apparent treatment efficacy; toxicity profile more complex; timing of co-interventions (baseline vs. post-randomization) must be determined carefully	Effect size may be diluted; false attribution of causality;
Other		
Prognosis	Poor prognosis participants more likely to be lost to follow up; unmeasurable effect (treatment response) introduces	Loss of follow up assessment can severely

Factors	May Bias Estimation of Response and Assessment of Tolerance	May Bias Effect Size
	severe discordance to study protocol execution; missing data may be informative and non-random	compromise validity of statistical inference

Because patients' health status are heterogeneous, a study protocol should describe carefully the target population (e.g. those with ECOG performance status <=1 and bone-only distant metastases), and then describe the eligibility of the participants to be recruited, with carefully justified inclusion and exclusion criteria. The sample of participants should allow appropriate generalizability of the trial results, but not be so open-ended as to lose sight of modest but meaningful efficacy, which may still be diluted by inherent limitations of symptom measurement methods or irreversible decline in the participant's health state. When designing and executing clinical trials in the palliative care population, it is also important to prepare for numerous obstacles in trial accrual, because many patients representative of the target population may be unable or unavailable to participate in the clinical trial. Hagen et al offer a taxonomy to categorize patients through the typical clinical trial screening process, accounting for the "total pool of potential study candidates" (an approximate target population), then those who consent to participate (total accrued), followed by those who successfully complete screening procedures, then those entered into the protocol intervention and finally those who complete all trial procedures and follow up evaluations.[17] The "number of patients needed to be in the initial study pool", or NNA, not only helps gauge generalizability of trial results, it can also inform other clinical trial endeavors on how they might refine accrual strategies. A thorough study protocol will aim to capture this valuable information.

Patient heterogeneity, even if controlled by restricting study entry based on pre-specified selection criteria, diminishes the precision of the efficacy measure and therefore the apparent effect size. One may increase the sample size to maintain power (accounting for relevant stratification variables). In addition, it could be valuable to pre-specify a multivariable model to adjust the effect estimate for the primary object in a randomized trial,[12] although phase 1 and phase 2 studies typically would not have sufficient sample sizes for multivariable adjustment. Multivariable analysis is especially useful if the outcome of interest has a non-linear distribution, such as in a binomial outcome (difference in proportion, i.e. logistic model) or in a time-dependent event outcome (i.e. proportional hazard model). This is because "the unadjusted estimate of the treatment effect is not correct if there is moderate heterogeneity of subjects, even with perfect balance of baseline characteristics across the treatment groups."[12] It is true that random allocation allows the estimated treatment effect to be attributed to the intervention, but other patient factors can also influence the probability of the outcome. This is especially true in a heterogeneous population where the risk of experiencing the outcome may vary, and therefore the absolute benefit from treatment may not be evenly spread across different risk groups. In other words, even though randomization is expected to balance baseline characteristics (known and unknown confounders) among the treatment arms, those patient- and tumor-related factors remain active in determining the outcome of interest. As discussed in the previous section, the multivariable model would have to be pre-specified to include clinically relevant and meaningful patient-, disease- and/or prior treatment-related factors, in order to provide confirmatory evidence of effect. A robust treatment effect would be stable with and without multivariable model adjustment.

Issues Related to Attrition and Missing Data

Among many challenges in designing and executing clinical trials involving palliative care patients, the most serious and perhaps under-appreciated problem is loss of follow up data, especially before the primary outcome is determined. Common reasons are:

1) a significant decline in health (patient unable to attend follow up assessment appointments or too fatigued to complete required questionnaires);
2) an intervening adverse event or coincidental illness when due for evaluation of study outcome (e.g. admission for hospital for unrelated thromboembolism, infection, failure to thrive);
3) migration (some patients move back to their hometowns for the anticipated final phase of illness);
4) death due to cancer progression, treatment-related adverse event or other conditions.

A common strategy to counter this is to inflate the sample size by an estimated proportion of expected attrition (an educated guess) for the target population. But this only addresses the power reduction due to attrition. The more serious and potentially critical problem is the bias resulting from "non-ignorable missingness", and, in the case of time-dependent event analysis (survival analysis), *informative censoring*. The problem of missing data is more serious if there is an important difference between the randomized arms that might become hidden, e.g. new treatment producing more toxicity and preventing follow up evaluation of both adverse events and the primary outcome.

Missing data is typically classified into three types:[12,18,19]

1) Missing completely at random: outcome data is missing independent of baseline or study factors and independent of the value of the outcome. For example, in a pain intervention trial, missing an assessment because of a pre-planned vacation out of town or a snow storm, but not because the patient has experienced a pain exacerbation or significant pain relief. The reason for missing data is unrelated to both the baseline patient status (disease status and intervention) and the outcome of interest (e.g. pain relief). In such a case, there is loss of precision of the analysis (an extra random error) but not added bias.
2) Missing at random: outcome data is missing and the probability of missing that outcome is independent of the outcome itself but may be related to a baseline factor. For example, missing outcome data is dependent on patient performance status (e.g. a higher proportion of ECOG 2 patients miss their follow up than ECOG 0 patients), but that missingness within each performance group remains independent of the outcome itself, as in (1) above.
3) Informative missingness or non-ignorable missingness: the observed outcome data has been selected in some way. For example, the primary follow up appointment is missed because of pain progression (or significant pain relief such that the patient decides to go on vacation), when pain is the primary outcome. If the outcome is to be analyzed using time-dependent method (i.e. survival analysis or proportional hazard modeling), that missingness still results in a biased analysis, even if the patient is to

be censored at the last follow up time point. This is because the analysis involving time-dependent censored data assumes that the censoring is non-informative, which requires that the censored individual to have the same risk to experience the primary outcome as those still remaining in the study. Unfortunately, in the clinical realities of managing palliative cancer patients, missing follow up is often due to deteriorating health status, usually a sign of failure of interventions, and the resulting missing data could well be informative. Ignoring this information in the evaluation of the interventions' effect could be a serious source of bias.

In the design and execution of a controlled clinical trial for a palliative cancer patient population, a thoroughly developed protocol would delineate operational procedures to determine the specific reasons for missing outcome data, develop strategies to avoid it (including choosing an efficacy outcome measure that best matches the health status of eligible patients), and pre-specify exploratory data analysis to characterize patterns of missingness relative to those participants that do have an observable primary outcome. The latter strategy is advisable especially when an appreciable proportion of accrued participants are missing. This could be done with techniques such as logistic regression (modeling missing or not missing by baseline characteristics at study entry) or recursive partitioning. Whatever the exploratory method, patterns of missingness should be reported to help readers understand the limitations of incomplete data.[12] Imputational methods add considerable complexity to the statistical analysis plan, and are mainly developed for missing at random. It is important to note that the investigator should not expect imputational methods to salvage a study with significant numbers lost to follow up or to fix informative or non-ignorable missingness problems.

Table 5. Considerations on clinical trial protocol design to reduce missing follow-up data

• Engage other health care professionals (especially palliative care physicians and home care nurses) in the design and execution of the trial
• Match palliative patient health status to efficacy measurement; healthier patients provide more reliable longitudinal and multi-modal follow up data; patients in poorer health are much more restricting with respect to extent and frequency of assessment
• Investigators tend to underestimate difficulties with multi-attribute questionnaires for patient reported outcomes (symptom-scales, health-related quality of life, and/or utility instruments); evaluate burden of efficacy measures a priori as part of a feasibility study
• Coincide trial follow up visit with other standard clinical visits (e.g. visits with another clinician, for pharmacy or diagnostic test)
• Define and defend choice of primary endpoint; plan a single and simple but validated endpoint for more advanced cancer patients
• Conduct a feasibility study or studies to evaluate logistics and protocol compliance [ref.36]
• Explore availability of standardized secondary data sources (e.g. electronic medical records or community care reports) for secondary outcome measures (e.g. toxicity & adverse events) based on strictly-defined data capture rules

A conservative approach to analyzing results with missing data is to assume all missing outcomes to be equivalent to treatment failure (or absence of a successful protocol-defined outcome) while maintaining an intention-to-treat analysis approach. However, in the setting of palliative care trials, where absence of follow up is as likely or more likely to be related to disease and symptom progression (i.e. failed primary outcome), the resulting effect observed is likely to be biased. In non-inferiority trials, the bias will tend towards concluding non-inferiority, and the generalizability of the results could be severely compromised. Sensitivity analysis could be done by limiting the estimate of treatment efficacy to only those participants with complete follow up evaluations for the primary endpoint, but the resulting effect size would still be difficult to generalize to the target population. In essence, it is important to *design* a clinical trial for the palliative care population with a strong appreciation of how participants may be lost to longitudinal follow up assessments. Strategies to *avoid* missing data should be delineated, especially regarding primary endpoint evaluation. Some suggestions are offered in Table 5.

Issues Related to Efficacy Measures

In appraising the potential clinical relevance and impact of an intervention trial in palliative care, an essential component is the primary efficacy measure, upon which the trial sample size and statistical inference will be based. However, as symptom measurement is itself a complex field of study, this chapter can offer only limited discussion and suggestions. The reader is referred to the reviews and recommendations by IMMPACT and National Cancer Institute of Canada (NCIC) Clinical Trials Group (CTG), among others, for further reading. [20-22]

- A clinical trial proposal must justify the choice of efficacy measure based on the study hypothesis and objectives, and at the same time being cognizant of its construct validity and reliability in the palliative care population. Because hypothesis testing or effect estimation necessarily narrows the efficacy to one or a few symptoms- or health-related measurements, the justification must also explain the compromises in the particular choice of efficacy measure (choosing one aspect of an outcome means that other aspects will be held secondary or not evaluated at all), and how the compromises would be acceptable and still provide sufficiently convincing information (in the form of effect size and its variability) at the end of the study. Patient reported outcome has become the standard approach to evaluating palliative endpoints (e.g. symptoms such as pain or shortness of breath; index symptom as chosen by patient; health related quality of life measures). The choice of palliative endpoint requires considerable discussion, as Brundage et al have noted discordance between symptom-based and health-related quality of life-based outcomes.[21]
- The proposal must also discuss and define minimal clinically important difference, which may be different in magnitude within one individual (before and after intervention) and between groups. [20]. Statistically significant change cannot be equated with clinically important change, which requires reasoned judgment based on prior evidence and pre-specification in the proposal. For example, a 0.5 point change in mean score on a 0-10 linear scale may be statistically significant in a trial

with large sample size, but its meaning is vague at best if the standard deviation is 2 or 3. The protocol must *justify* the choice of efficacy as to represent both clinically meaningful change to the individual after intervention (anchor-based method), and to discriminate between comparison groups (e.g. 10% difference in response rate between arms). Alternatively, a distribution-based approach can also be declared and justified.[20]

- If multiple analyses (more the one primary endpoint) or a composite primary endpoint is planned, clarification and justification of the clinical decision rule should be discussed (see "Attention to detail…" section above). For a composite primary endpoint (an example would be pain combined with analgesic use), each component must be well-defined, justified, and assessed with equal rigor. Moye explains four applicable principles (prospective deployment or specification, coherence, precision and disclosure) that are easy to comprehend, and Turk et al provide a concise description and discussion. [1,22].

A key point of discussion here for the motivated clinician researcher is that there will always be *uncertainty* with respect to whether the experimental intervention will be equal to or have "significant" improvement over the standard of care, which is of course the purpose of doing the study in the first place. The investigator must consider the nature of the equipoise being offered to the participants through the study, without subconsciously wanting to "fix" the outcome of the trial in favor of the investigator's preference (e.g. by searching for statistical significance without considering clinical importance). The investigator should also resist the fear of "picking the wrong measure" for primary endpoint in the proposal, as well as the temptation to declare a secondary endpoint as conclusive. Choosing an appropriate efficacy measure is best guided by careful considerations and deliberations of what would be clinically relevant and beneficial to the target patient population in question. An understanding of the clinical needs or gap is essential, keeping in mind that the most efficient measurement may not be clinically meaningful necessarily.

Methodological Issues in Palliative Radiotherapy Trials

Palliative effects of therapeutic radiation in advanced cancers were recognized early in the discovery of X-rays, particularly for treatment of cancer pain. Much of the effort in palliative RT trials in the past decades focused on relief of cancer-induced bone pain, either through localized [23] or wide-field RT [24] and systemic radionuclides.[25,26] RT trials of brain metastases have been numerous, but most are concerned with survival as the primary endpoint [27,28], with very few designed with patient-reported outcomes as the primary endpoint.[29,30] The palliative effects of RT in non-small cell lung cancer were evaluated in an NCIC CTG trial SC 15.[31] Of note, in almost all of these palliative RT trials, comparisons were made with active controls (a different RT schedule or RT in combination with a systemic agent).

Issues concerning palliative RT trials begin with the nature of the symptom(s) and conditions to be considered for palliation, the perceived clinical needs of the population, the

therapeutic advantage of radiation towards the symptom(s) in question, and the availability of other treatments that may be co-interventions, such as opioids, corticosteroids, bisphosphonates, palliative surgery, and systemic cytotoxic or cytostatic agents. In addition to patient heterogeneity issues discussed earlier, symptom heterogeneity may strongly influence the theoretical benefits (therapeutic advantage) of RT. Some of these factors are outlined in Table 6.

Table 6. Factors contributing to heterogeneity of cancer symptoms

Factors	Levels of Symptom Manifestations
Anatomic focality	Localized or systemic source of symptom (multifocal symptoms can occur concurrently or sequentially, and temporality may change as disease progresses)
Pathophysiologic	• Cellular, microscopic reactions e.g. nerve infiltration, inflammatory cytokines • Mass, macroscopic effects e.g. obstruction, destruction, compression, neovascularization effects • Organ compromise e.g. pathologic fracture, increased intracranial pressure, spinal cord/cauda equina compression
Intensity/Severity	Mild, moderate, severe, accompanied by varying degrees of impairment
Acuity-chronicity	• Acute symptom episode that fades or drops to stable level • Intermittent exacerbation of chronic symptom • Stable but progressing chronic symptom
Clustering	Index symptom (e.g. pain or dyspnea) associated with others such as fatigue, anorexia, drowsiness, anxiety, etc. Targeting one symptom alone while another symptom worsens may not produce a net palliative benefit
Adaptation	Activities (e.g. eating, sitting, standing, walking, talking) are altered to varying degrees to balance comfort preference and impairment/disability; symptom experience is partly dependent of patient environment

The challenge with designing trials incorporating new palliative RT approaches is to balance the theoretical advantage of radiotherapy treatment, presumably through immune-mediated responses (acute) or neoplastic cytostasis/cytoreduction (latent), against its adverse effects, mainly fatigue and acute inflammatory reactions (the –itis reactions). Because of the compromised health status common to palliative patients, adverse effects that would otherwise be well-tolerated in healthier cancer patients may become profound in the palliative patient. One is cognizant that grading of performance status or adverse effects on a linear scale does not mean the clinical effect is indeed linear. For example, a 1-point increase in fatigue from grade 1 to grade 2 means some loss of work ability, but from grade 2 to grade 3 mean potential loss of independence. The same 1-point increase may carry profound difference in patient experience and outlook. As a result, the therapeutic advantage of palliative RT is not easily conceptualized and thereby exploited to tilt the risk:benefit ratio towards benefit. Strictly concerning symptom relief, manipulation of dose-schedules (fractionation and dose-intensity) has produced limited added benefits, partly because palliative RT can be repeated over time, as long as the total cumulative dose to a given volume is within reasonable tolerance limits.

Recent advances in RT planning and delivery have provided clinical researchers with experience in the delivery of conformal, normal-tissue sparing RT, most notably stereotactic radiosurgery or stereotactic RT. This has allowed the delivery of modest-to-high dose ablative

radiation with apparent safety even in the presence of multiple target volumes in more than one organ system in metastatic cancer patients. Concerning clinical trial design issues, such new approaches, when applied in the palliative cancer population for purpose of *symptom relief*, will encounter many of the aforementioned methodological challenges.

A related and more intriguing clinical trial design issue is in the application of multi-focal ablative RT for the purpose of *disease control* in patients with metastatic disease, apart from or in addition to the control of symptoms. The notion of applying RT in such a manner, namely in patients with "oligometastases", is akin to the development of systemic therapies, and the diagnostic-imaging based, composite endpoint measurement approach, namely RECIST, comes to mind as a potential way to measure efficacy.[32] One thing for certain, such a trial will have to *pre-specify either* a symptom control effect *or* disease control as the primary study objective. The enthusiastic investigator would be tempted to deploy a composite endpoint based on both types of effect, or, alternatively, pre-specify a multiple-endpoint analysis strategy. As discussed in the preceding sections, such a construct would encounter heterogeneity problems with patient selection, need stringent demands for protocol execution, and demand a complex and well-reasoned statistical analysis plan.

From a health care perspective, justification of a palliative RT intervention will also have to give considerations to its contribution supplementary to other medications or interventions in question, if it is proposed "to treat a disease for which there are already a panoply of medications":

> "In this situation where the standard of care is 'acceptable', the new medication must demonstrate substantial efficacy in order to justify its use in a clinical environment in which practitioners are comfortable with the risk-benefit balance of the standard treatments. If, in addition, the medication that is being proposed is costly and/or has a new and serious adverse event profile, the efficacy threshold for the medication must be even higher in order to preserve the risk-benefit balance." [1]

Essential to a RT intervention protocol is the technical detail involved in the planning and delivery of radiation, with varying degrees of complexity in tumor-anatomic target definition (accuracy of delineation and reproducibility on treatment), dosimetric parameters with respect to target and normal tissue exposure, and treatment quality assurance. In a population where uncontrolled symptoms necessitate evaluation by the radiation oncologist, clinical trial logistics have to meet clinical expectations of time-to-treatment considerations ("waiting time"), especially if the symptoms are distressing or the patient is scheduled to have concurrent or sequential systemic therapy. The complexity of the radiotherapy invention must be feasible, including but not limited to its logistics in the clinical cancer environment and the consistency of its delivery across different treatment centres.

Conclusion

Clinical trials and indeed most clinical research endeavors represent sample-based research whose ultimate goal is to infer experimental or observational results from the sample to the population to promote health and quality of life. This process demands a thoughtful, disciplined approach to the design and execution of the study protocol. An understanding of

the principles of statistical inference is fundamental to both the design and the valid interpretation of the research program. The clinical investigator would be prudent to collaborate with experienced investigators from different disciplines, keeping in mind that the value of one innovation is relative to another, and that the true benefit is reaped not with the publication of a completed trial, but when the community at large is at a better state of health.

Conflict of Interest

The author declares no conflict of interest.

References

[1] Moye LA. Statistical reasoning in medicine: the intuitive P-value primer. 2nd ed. New York: *Springer Science* + Business Media; 2006.

[2] Fletcher RH, Fletcher SW, Fletcher GS. Clinical epidemiology: the essentials. 5th ed. Philadelphia:Lippincott Williams & Wilkins.

[3] Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJW. Reporting of noninferiority and equivalence randomized trials: an extension of the CONSORT statement. *JAMA*. 2006;293:1152-60.

[4] Ramsey FL, Schafer DW. The statistical sleuth: a course in methods of data analysis. 3rd ed. Boston: Brooks/Cole Cengage Learning; 2013.

[5] Goodman SN. Toward evidence-based medical statistics: 1. the P value fallacy. *Ann. Intern. Med.* 1999;130:995-1004.

[6] Goodman SN. A dirty dozen: twelve P-value misconceptions. *Semin. Hematol*. 2008;45:135-40.

[7] Silva-Aycaguer LC, Suarez-Gil P, Ferna. The null hypothesis significance test in health sciences research (1995-2006): statistical analysis and interpretation. *BMC Med. Res. Method*. 2010;10:44.

[8] Austin P, Mamdani MM, Juurlink DN, Hux JE. Testing multiple statistical hypotheses resulted in spurious associations: a study of astrological signs and health. *J. Clin. Epidemiol.* 2006;59:964-9.

[9] Cao J, Zhang S. Multiple comparison procedures. JAMA. 2014;312(5):543-4.

[10] Moye L. Multiple analyses in clinical trials: fundamentals for investigators. New York: Springer-Verlag; 2003.

[11] Sribney B. What are some of the problems with stepwise regression? [Internet] College Station:StataCorp LP; 1998 [updated July 2011; cited 2014 Oct 1]. Available from: http://www.stata.com/support/faqs/statistics/stepwise-regression-problems/

[12] Harrell FE. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. New York: Springer-Verlag; 2001.

[13] Austin P, Tu JV. Automated variable selection methods for logistic regression produced unstable models for predicting acute myocardial infarction mortality. *J. Clin. Epidemiol.* 2004;57:1138-46.

[14] Sun P, Shook TL, Kay GL. Inappropriate use of bivariable analysis to screen risk factors for use in multivariable analysis. *J. Clin. Epidemiol.* 1996;49(8):907-16.
[15] Freemantle N. Interpreting the results of secondary endpoints and subgroup analyses in clinical trials: should we lock the crazy aunt in the attic? *BMJ*. 2001;322:989-91.
[16] Moye L. Random research. *Circulation.* 2001;103:3150-3.
[17] Hagen NA, Wu JS, Stiles CR. A proposed taxonomy of terms to guide the clinical trial recruitment process. *J. Pain Sympt. Manage*. 2010;40(1):102-10.
[18] Van Belle G. Statistical rules of thumb. 2nd ed. New Jersey: Wiley; 2008.
[19] Allison PD. Missing data. Thousand Oaks: Sage Publications; 2001.
[20] Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J. Pain.* 2008;9(2):105-21.
[21] Brundage M, Bezjak A, Tu D, Palmer M, Pater J. NCIC Clinical Trials Group experience of employing patient-reported outcomes in clinical trials: an illustrative study in a palliative setting. Expert Rev. Pharmacoeconomics *Outcomes Res.* 2008;8(3):243-53.
[22] Turk DC, Dworkin RH, McDermott MP et al. Analyzing multiple endpoints in clinical trials of pain treatments: IMMPACT recommendations. *Pain.* 2008;139:485-93.
[23] Dennis K, Makhani L, Zeng L, Lam H, Chow E. Single fraction conventional external beam radiation therapy for bone metastases: a systematic review of randomized controlled trials. *Radiother. Oncol*. 2013;106(1):5-14.
[24] Salazar OM, Sandhu T, da Motta NW, et al. Fractionated half-body irradiation (HBI) for the rapid palliation of widespread symptomatic metastatic bone disease: a randomized phase III trial of the International Atomic Energy Agency (IAEA). *Int. J. Radiat. Oncol. Biol. Phys.* 2001;50(3):765-75.
[25] Christensen MH, Petersen LJ. Radionuclide treatment of painful bone metastases in patients with bresat cancer: a systematic review. *Cancer Treat Rev.* 2012;38(2):164-71.
[26] Bauman G, Charette M, Reid R, Sathya J. Radiopharmaceuticals for the palliation of painful bone metastasis - a systematic review. *Radiother. Oncol*. 2005;75(3):258-70.
[27] Tsao MN, Lloyd N, Wong RK et al. Whole brain radiotherapy for the treatment of newly diagnosed multiple brain metastases. *Cochrane Database Syst. Rev.* 2012;4:CD003869.
[28] Patil CG, Pricola K, Sarmiento JM, Garg SK, Bryant A, Black KL. Whole brain radiation therapy (WBRT) alone versus WBRT and radiosurgery for the treatment of brain metastases. *Cochrane Database Syst. Rev.* 2012;9:CD006121.pub3.
[29] Kocher M, Soffietti R, Abacioglu U, et al. Adjuvant whole-brain radiotherapy versus observation after radiosurgery or surgical resection of one to three cerebral metasetases: rsults of the EORTC 22952-26001 study. *J. Clin. Oncol.* 2011;29(2):134-41.
[30] Langley RE, Stephens RJ, Nankivell M, et al. Interim dta from the Medical Research Council QUARTZ trial: dose whole brain radiotherapy affect the survival and quality of life of patients with brain metastases from non-small cell lung cancer? *Clin. Oncol.* (Royal Coll). 2013;25:e23-e30.
[31] Bezjak A, Dixon P, Brundage M, et al. Randomized phase III trial of single versus fractionated thoracic radiation in the palliation of patients with lung cancer (NCIC CTG SC.15). *Int. J. Radiat. Oncol. Biol. Phys.* 2002;54(3):719-28.

[32] Eisenhauer EA, Therasse P, Boggaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur. J. Cancer.* 2009;45:228-47.
[33] Le Tourneau C, Lee JJ, Liu LL. Dose escalation methods in phase 1 cancer clinical trials. *J. Natl. Cancer Inst.* 2009;101:708-720.
[34] Simon R. Optimal two-stage designs for phase II clinical trials. Cont Clin Trials 1989;10:1-10.
[35] Rubinstein LV, Korn EL, Freidlin B, Hunsberger S, Ivy SP, Smith MA. Design issues of randomized phase II trials and a proposal for phase II screening trials. *J. Clin. Oncol.* 2005;23(28):7199-206.
[36] Moore CG, Carter RE, Nietert PJ, Stewart PW. Recommendations for planning pilot studies in clinical and translational research. *Clin. Transl. Sci*. 2011;4(5):332-7.