Published in final edited form as:

J Vasc Interv Radiol. 2013 August ; 24(8): 1157–1164. doi:10.1016/j.jvir.2013.04.024.

Cancer Concepts and Principles: Primer for the Interventional Oncologist-Part I

Ryan Hickey, MD¹, Michael Vouche, MD¹, Daniel Sze, MD², Elias Hohlastos, MD¹, Jeremy Collins, MD¹, Todd Schirmang, MD¹, Khairuddin Memon, MD¹, Robert K Ryu, MD¹, Kent Sato, MD¹, Richard Chen, DO¹, Ramona Gupta, MD¹, Scott Resnick, MD¹, James Carr, MD¹, Howard Chrisman, MD¹, Al Nemcek, MD¹, Robert Vogelzang, MD¹, Robert J Lewandowski, MD¹, and Riad Salem, MD, MBA¹

¹Department of Radiology, Division of Interventional Oncology, Northwestern University, Chicago IL

²Department of Radiology, Stanford University, Palo Alto, CA

Keywords

interventional oncology; cancer; outcomes; response assessment; chemotherapy; interventional radiology

INTRODUCTION

A sophisticated understanding of the rapidly changing field of oncology, including a broad knowledge of oncologic disease and the therapies available to treat them, is fundamental to the interventional radiologist providing oncologic therapies and necessary to affirm interventional oncology as one of the four pillars of cancer care alongside medical, surgical and radiation oncology. Historically, interventional oncology procedures were first performed for palliative or exceptional cases. Now supported by robust evidence, this discipline is now gaining major importance in the management of cancer and is becoming incorporated in therapeutic algorithms as first line therapies. An interventional oncologist must understand the intricacies of clinical trials not only to critically assess evidence for new and evolving treatments, but also to design and carry out the clinical studies that will continue to define the role of interventional oncology. Just as fluency of expression and understanding in a language requires more than simple familiarity with its vocabulary, assuming a confident and informed role as an interventional oncologist requires proficiency with the tools and metrics of oncology.

© 2013 The Society of Interventional Radiology. Published by Elsevier Inc. All rights reserved.

Corresponding Author: Riad Salem, MD, MBA, Professor of Radiology, Medicine and Surgery, Vice-Chairman, Image-Guided Therap. Chief, Vascular and Interventional Radiology, Director, Division of Interventional Oncology, Department of Radiology, Northwestern University, 676 N. St. Clair, Suite 800, Chicago, Illinois 60611 USA, r-salem@northwestern.edu.

Conflict of Interest: Dan Sze reports the following disclosures: scientific advisory boards (Jenerex Biotherapeutics, SureFire Medical, Treus Medical, RadGuard Medical, Lunar Design); speaker's bureau (WL Gore), consultant (BTG, Sirtex). Robert Lewandowski reports the following disclosures: scientific advisory board (Surefire, Nordion). Riad Salem reports the following disclosures: scientific advisory board and consultant (BMS, Abbott, Bayer/Onyx, NCCN, Merit, BTG, Nordion, Sirtex, Boston Scientific). None of the other authors report a conflict of interest.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The first part of this review intends to provide a concise overview of the fundamentals of oncologic clinical trials, including trial design, methods for assessing therapeutic response, common statistical analyses, and the levels of evidence provided by the clinical trials. The second part of this review will focus on methods of tumor characterization, the principles of medical, surgical, radiation and interventional oncology, and the current treatment paradigms, including the levels of evidence behind the treatments, for cancers most commonly encountered in interventional oncology. Terms of particular importance will be underlined in order to highlight their importance in the interventional oncologist's lexicon.

CLINICAL TRIAL DESIGN

Cancer treatment trials are designed to evaluate new therapies and are divided into different phases of evaluation for anticancer agents according to the United State Food and Drug Administration (FDA) (1). Phase I trials evaluate the safety and toxicity of a compound and are typically performed in a dose-escalation format. Progressively higher doses are administered to determine the maximum tolerated dose and dose-related toxicities in order to establish a safe dose for the following phase of the trial. Phase II trials focus on determining the efficacy of an agent at its maximum tolerated dose. Because response rates and duration of response are the typical endpoints, patients must have measurable disease at baseline which, according to most criteria, must be at least 1 cm (2, 3). Phase III clinical trials are conducted once the maximum tolerated dose of an agent has been tested in phase II study and has demonstrated a reasonable response. Phase III trials typically compare the test agent to standard of care therapy. Although phase III trials are the definitive studies to demonstrate superiority of a new therapy over the previous standard, many new cancer agents are FDA approved after successful phase II studies. Phase IV clinical trials, often referred to as post-marketing or post-approval studies, are conducted to gather more information about the real-life clinical applicability of a therapy after it has been approved in order to evaluate its clinical implementation, expand labeling and approval, or meet requirements established by a regulatory authority.

Randomized, controlled trials are the gold standard of clinical studies. Comparative trials (non-randomized) are vulnerable to various sources of bias and therefore study conclusions must be interpreted with caution. Propensity scores and match-pairing can help minimize the effects of bias on this type of study. While randomization produces comparable groups by eliminating selection and allocation biases, on average, one should confirm that factors affecting prognosis and outcome are evenly balanced between the control and intervention groups. The lack of selection and allocation bias and reliable comparability of randomized groups is also required for accurate use and interpretation of tests of statistical significance (4).

Blinding is another important factor in determining the validity of clinical trial results and conclusions. In unblinded studies, both the participant and the investigator know the intervention to which the participant has been assigned. While this model more accurately reflects clinical practice, it opens the results to bias. For example, participant reporting of side effects or investigator treatments may be affected and change the course of therapy and results. However, certain treatments do not lend themselves to complete blinding. In a single-blind study, only the investigator knows to which group the participant has been assigned. It eliminates participant reporting bias but does not eliminate bias with regard to investigator behavior and data assessment. Double-blind clinical trials eliminate investigator knowledge and therefore investigator bias. Unfortunately, knowledge of specific drug side-effects often prevents the complete blinding of treatment arm. As an example, sorafenib produces specific toxicities that would be clearly recognized by investigators during the

management process. In triple-blind studies, the individuals evaluating data and results are not aware of the participant treatment groups (5).

Intention-to-treat analysis is a critical component of clinical trials. It indicates that all patients assigned to a study arm at the time of randomization are analyzed regardless of subsequent events, such as noncompliance with treatment, receiving no treatment, crossover, loss to follow-up and dropping out. It eliminates bias that might overestimate the clinical benefit of the therapy and ensures that “real-world” effects of the treatment or control are reflected in data analysis. In contrast, in a per-protocol analysis, only patients completing the clinical trial according to the trial protocol are evaluated (6).

ASSESSMENT OF RESPONSE

The outcomes of oncologic therapies can be measured according to several different criteria, each of which reflects a different clinical value that can help determine the applicability of statistically significant results.

Tumor Response Assessment

The World Health Organization (WHO) criteria for reporting results of cancer treatment were the first standardized approach to report degrees of tumor response, recurrence and disease-free interval (2). In a bi-dimensional approach, maximal tumor cross-sectional area is approximated by measuring the mathematical product of the longest diameter and longest perpendicular of each tumor present at the baseline assessment. These products are then summed to obtain the value used in assessing response. Response assessment must occur at intervals of no less than four weeks. Complete response is defined as disappearance of all known disease, while a partial response is defined as 50% decrease in the sum of the diameter products. Disease progression refers to >25% increase in the size of one or more tumors, or the appearance of a new disease. Stable disease indicates a lack of partial response or disease progression.

In 2000, a new model for evaluating response in solid tumors was proposed in order to address areas of conflict and inconsistency that had arisen in the use of WHO criteria. The Response Evaluation Criteria in Solid Tumor (RECIST) group proposed a model by which a combined, uni-dimensional assessment of all existing lesions is used to extrapolate an overall response to treatment. Solid tumors are defined as target and non-target lesions, with the target lesions defined as those to be measured (7). The RECIST criteria were updated in 2009 as the RECIST 1.1 criteria, in order to address issues that had arisen in the use of the RECIST 1.0 criteria (3). In RECIST 1.1, the number of lesions required to assess tumor burden was reduced to a maximum of five total lesions with no more than two lesions per organ. Methods for assessing pathologic lymph nodes were included. Progressive disease was updated to reflect not only a relative increase in target disease, but also to establish a minimum absolute increase in the setting of small lesions. Recommendations for optimal anatomical assessment on imaging were also included.

The WHO and RECIST criteria have limitations in certain types of cancer. By relying upon changes in tumor dimension as a measure of response, the WHO and RECIST criteria do not take into account tumor necrosis that occurs without a reduction in the overall tumor size. Indeed, anatomic tumor measurements using the RECIST criteria do not reflect the extent of tumor necrosis resulting from locoregional interventional radiology procedures or molecular targeted therapies (8). The European Association for the Study of Liver (EASL) recommended modifications to the bi-dimensional WHO response criteria for HCC to estimate the reduction in viable tumor volume, as evidenced by enhancing areas on CT (9). More recently, the modified RECIST criteria (mRECIST) were proposed and are based on

measurements of tumor viability as indicated by arterial phase enhancement. mRECIST also includes guidelines for assessing pleural effusions and ascites, porta hepatis lymph nodes, portal vein thrombosis and the development of new lesions (10). Most importantly, EASL and mRECIST criteria both assess tumor enhancement (assumed to represent tumor viability), with EASL criteria based on the product of two diameters of enhancing tumor whereas mRECIST relies on measurement of the longest axis of enhancing tumor (Table 1).

With regard to HCC, additional imaging techniques have been described for response assessment but have yet to be fully validated or accepted. Changes in the apparent diffusion coefficient in diffusion-weighted MR imaging (DWI) have been described to assess response in HCC following percutaneous ablation, transarterial chemoembolization and transarterial radioembolization, with increases in the ADC value of the tumor being associated with tumor response (11–14). However, at explantation and pathologic analysis, contrast-enhanced MRI has been shown to more significantly correlate with histopathologic findings of tumor necrosis compared to DWI (15), and tumor contrast-enhancement remains the measurable entity for response assessment according EASL and mRECIST criteria.

Anatomic imaging using changes in tumor size and enhancement for response assessment do have limitations, particularly following therapies with cytostatic rather than cytotoxic effects. Criteria based on functional imaging and metabolic activity with 18-fluorodeoxyglucose positron emission tomography (18F-FDG-PET) have been proposed, referred to as PERCIST for PET Response Criteria in Solid Tumors, and may be incorporated in response assessment of certain solid tumors as well as metastases (16).

Clinical Response Assessment

a) Efficacy versus Effectiveness—When discussing response to therapy, the difference between efficacy and effectiveness must be clarified. Efficacy refers to the ability of the therapy to bring about its intended effect in an ideal setting. Effectiveness, on the other hand, refers to the ability of a therapy to bring about its intended effect in real-life circumstances. A therapy may have significant efficacy in terms of tumor response or patient survival when the therapy is delivered as intended, but a high toxicity profile may preclude administration of all of the intended doses or reduce patient compliance, thereby limiting the effectiveness of the therapy (17).

b) Overall Response Rate—The overall response rate (ORR) refers to the proportion of patients in a study with a measurable reduction in tumor size, either by radiologic assessment or physical examination. Response duration is measured from the time of response until the time at which progression is identified. The Food and Drug Administration (FDA) generally defines ORR as the number of patients with complete response (CR) and partial response (PR). This typically does not include stable disease (SD), which indicates either a lack of observable progression or tumor shrinkage less than that which is considered a response. Stable disease (SD) is also not included in ORR as it may be reflective of the tumor biology, such as slow growth seen in the setting of neuroendocrine tumors or small hepatocellular carcinomas, rather than the therapeutic effects of a drug or intervention. In general, radiological objective response rate is often a surrogate endpoint in phase II trials, whereas survival remains a key endpoint in phase III trials (18). Disease control rate (DCR) is represented by the sum of CR+PR+SD. While DCR is often a reported outcome, this parameter should be interpreted with caution. When cancer treatment is undertaken, the assumption is that active tumor growth will be observed at subsequent scans and that achieving SD is felt to reflect anti-tumoral effect. This may not be that case with tumors where there is slow growth and extended natural history of disease (ex: early stage

hepatocellular carcinoma, neuroendocrine tumors). In such cases, although a very high DCR will be reported, much of it may be attributable to slow natural history.

c) Overall Survival—Overall survival (OS) remains the gold standard endpoint for cancer research and trials, as well as for measuring treatment effectiveness by the United States Food and Drug Administration (FDA) and the European Medicines Agency (EMA) (19). Overall survival is defined as the period of time from randomization to death. As long as quality of life is not diminished, overall survival reflects the most valuable benefit (survival) to patients. It is an important clinical measure as it can identify therapies for which there is high treatment-related mortality despite good or better tumor control. However, overall survival as an endpoint and indicator of investigational therapy effectiveness can be confounded by the use of subsequent alternative treatments, which are already known to be effective, after disease progression or discontinuation of the study drug (18) (20).

d) Time-to-Progression and Progression-Free Survival—Time-to-progression (TTP) is defined as the time from randomization to the time of objective disease progression. In measuring TTP, death occurring without evidence of tumor progression are censored, or not included, in data analyses.

Progression-free survival (PFS) refers to the time from randomization to the time of objective disease progression or death. PFS captures deaths that may be related to adverse effects of therapy. By evaluating only those patients who survive a potentially toxic therapy and ignoring those who die without progression, as in a TTP analysis, a therapy might be incorrectly characterized as being more effective. Because TTP analyses censor deaths without progression, a TTP analysis may indicate a longer time without progression compared to a PFS analysis for the same study population. In fact, as opposed to PFS, it is possible for TTP to be longer than overall survival (death without progression) (18).

Benefits and Limitations to Clinical Response Assessment

Although OS is the gold standard endpoint for cancer trials, surrogate markers of overall survival, such as TTP and PFS, can be important and sometimes necessary endpoints in terms of practicality. Compared to assessments of OS, TTP and PFS studies have earlier trial endpoints and are less expensive to conduct. By including progression, which is a usual precursor to death, PFS provides a survival assessment with an earlier endpoint that may correlate with survival (21, 22). PFS endpoints, for example, are frequent in breast cancer and neuroendocrine studies due to prolonged survival times and a need for measurable endpoints in a practical timeframe.

PFS and TTP analyses can provide information on the effectiveness of a particular therapy without the confounders of cross-over and subsequent therapy. Cross-over and subsequent-therapy effects can mask the tumor control effects of a particular therapy since an OS analysis includes not only the effects of the therapy under investigation, but also the effects of all subsequent lines of therapy. Given that disease progression is often the time at which patients are started on a subsequent therapy or cross over to another treatment arm, a TTP or PFS analysis would reflect the endpoint and eliminate the confounder (22, 23). As surrogate markers of treatment efficacy in clinical trials, PFS are often more appropriate when evaluating efficacy of second or third line therapies (ex: Y90 for recurrent liver metastases). On the other hand, TTP maybe more useful for assessing efficacy of first line therapies (ex: RFA for small HCC discovered on routine imaging), since recurrence is more likely to occur after 1st line therapies, while death will often occur without progression after multiple lines of therapies.

Whereas OS analysis have the unequivocal endpoint of patient death that reflects both safety and efficacy of a treatment, PFS and TTP analyses can be confounded by the frequency in which, or interval of time between which patients are assessed for progression, resulting in a time-dependent assessment bias. A treatment that causes patients to undergo clinical evaluation or imaging more frequently can cause disease progression to be identified sooner, while an extended interval between assessments could result in failure to detect an actual difference in disease progression. TTP and PFS analyses should maintain similar time intervals between assessments in the treatment and control arms in order to minimize this bias. Short-interval follow up, such as imaging obtained at 6 to 8 week intervals, can help to accurately measure TTP and PFS (22).

For diseases in which death is most often caused by disease progression, one should expect values of TTP and PFS to be equivalent. A PFS shorter than TTP in this circumstance would indicate significant treatment toxicity. For patients whose malignancies are often associated with a particular comorbidity, such as HCC and cirrhosis, TTP as opposed to PFS, is a valuable study endpoint for phase II studies to help distinguish patients who reached a study endpoint due to the malignancy under study as opposed to the comorbid condition (24).

When measuring response to therapies that tend to be cytostatic rather than cytotoxic, survival-based analyses such as PFS and OS are particularly valuable since effective cytostatic therapies may not demonstrate a measurable change in tumor dimensions necessary for an objective tumor response. This is especially true in the assessment of molecularly-targeted therapies, which tend to have predominantly cytostatic effects on tumor progression. For example, the SHARP and Asia-Pacific trials demonstrated a significant prolongation of survival for patients with HCC treated with sorafenib despite objective tumor response rates by RECIST of only 2% and 3.3%, respectively (25, 26).

Disease-free survival is a similar metric of response to PFS but implies that patients were disease-free at baseline evaluation. It is most often used when assessing adjuvant therapy following definitive surgery or radiotherapy. Recurrence-free survival is a similar endpoint, although more specific to the site of a curative therapy, such as following resection or ablation. When the presence of metastases is a reliable indicator of death, time-to-distant metastases is an additional endpoint that may be used as a surrogate for overall survival (5).

BASIC STATISTICS IN ONCOLOGY

Time-to-endpoint analyses (Kaplan-Meier)

In a Kaplan-Meier plot (time-to-endpoint analysis), occasionally referred to as survival plots, the number of patients who are progression-free and alive (in a PFS analysis), or alive (OS analysis), at the end of a specified interval are divided by the total number of patients at the beginning of the interval. The X-axis represents time and the Y-axis represents the cumulative proportion of patients free from disease progression or death. If a patient has been censored from data analysis, they are not included in the calculation of surviving study patients. Patients may be censored because they drop out of the trial or are lost to follow-up before the endpoint of disease progression or death has occurred. They are included in the calculation at time intervals preceding the time of censoring, but will not be included in subsequent calculations at or after the time of censoring. Censored patients are indicated by dots or vertical marks on the Kaplan-Meier curve (5).

Hazard/Relative Risk Ratio, Absolute Risk Reduction, and Number Needed to Treat

The hazard ratio (HR), expressed as a rate or percentage, is derived from Kaplan-Meier calculations and indicates the probability that a patient will experience a hazardous event during a specified time interval. It is also a way of quantifying the difference of two Kaplan-

Meier curves. The hazard ratio differs from the relative risk ratio in that the hazard ratio refers to probability of an event over a specific time interval, whereas the relative risk ratio refers to probability over the entire course of a study and is directly related to the study endpoint. The lower the hazard ratio, the more favorable the treatment arm is compared to the control or standard therapy arm. As opposed to describing a median (one time point), the HR expresses the benefit (or lack thereof) of a treatment for all time points.

The relative risk ratio indicates whether two characteristics are significantly related, such as a risk factor and an event. It is a ratio of the event rate in the intervention arm compared to the event rate in the control arm. The absolute risk reduction (ARR) is the absolute value of the difference of events occurring between the intervention and control arms. It can be thought of as the number of people per 1000 or 10,000, to whom the event does not occur. The relative risk reduction (RRR) is the percent reduction in risk. The number needed to treat is the reciprocal ($1/ARR$) and indicates the number of people needed to treat in order to prevent one event (27).

Correlation

For analyzing correlation between interval values, the Pearson product moment correlation coefficient is often used, while Spearman's rank correlation coefficient is generally used for ordinal values or values in which there are significant outliers. The correlation coefficient lies between -1.00 and $+1.00$, with -1 indicating a perfect negative linear correlation and $+1$ indicating a perfect positive linear correlation. In general, absolute values of the correlation coefficient from 0 to 0.25 indicate very little or no association, 0.25 to 0.50 indicate a fair association, 0.50 to 0.75 indicate moderate association, and 0.75 to 1.00 indicate a very high degree of association. The coefficient of determination (r^2) indicates the percentage of variability that can be predicted by knowing one of the two values. That is, knowing one of the correlates, the variance of the second correlate would be reduced by r^2 (27).

Confidence Intervals

Confidence intervals, or confidence limits, are indicators of the variability and reliability of an estimate and indicates the probability that the interval provided will contain the true parameter value. For example, a 95% confidence interval indicates that the likelihood that the true value is outside of the interval is less than 5% (5).

The interquartile range (IQR) is another indicator of statistical dispersion and demonstrates the difference between the third (upper) and first (lower) quartiles. It indicates the range comprising 50% of the data (28).

Error and Power

Clinical studies generally test for a difference between the rates of a particular event in two groups. The null hypothesis states that there is no difference, and the study is performed in order to accept or reject the null hypothesis. The null hypothesis is assumed to be true until proven otherwise. When, simply by chance, a sample is not truly representative of the larger population, a study may find a difference between two interventions, even though such a difference does not actually exist in the larger population. This false positive finding is referred to as type I error. The significance level, denoted by alpha (α), indicates the probability that the sample studied provides artificial evidence of a difference. In other words, α is the probability of having a type I error event (29). Caution should be exercised in cases of multiple comparisons, since a significant association may be observed by pure chance. There are several methods to correct for this type I error, the simplest being Bonferroni methodology (observed p value is multiplied by the number of tests). When a difference in outcomes between two interventions is small, it may not be detected and may

result in erroneously stating that there is no difference, when in fact there is a difference. This is referred to as type II error. The probability of a type II error event is denoted by β (). The power of a study refers to the probability that a type II error will not occur, equal to $(1 - \beta)$. A study of sufficient power will be able to demonstrate a difference when a true difference exists (30).

Since the null hypothesis and alternative hypothesis do not specify a direction of deviation for differences (e.g., positive/negative, increase/decrease) in the populations being studied, a two-tailed test is performed to consider both sides of the null hypothesis distribution, splitting between two tails. A directional hypothesis considers only one direction and therefore one tail, since the other direction is not a consideration for the alternative hypothesis. In such a one-tailed test, β is evaluated with regard to one tail. A one-tailed test involves a judgment or expectation of outcome prior to performing the study. In general, two-tailed tests should be performed to evaluate for differences in either direction, unless a large difference in one direction would be regarded as equivalent to finding no difference at all (31).

LEVELS OF EVIDENCE AND STRENGTH OF RECOMMENDATIONS

There are a number of different models used to indicate the strength of evidence acquired from clinical trials, as well as the strength of a treatment recommendation based on such evidence. Two of the most common systems used in the United States for oncology come from the National Cancer Institute (NCI) and the National Comprehensive Cancer Network (NCCN). For a given cancer therapy, the NCI offers a ranking of the strength of evidence according to an alphanumeric scale, which reflects both the strength of the study design and the strength of the study endpoints. Together the two rankings indicate the overall level of evidence for a specific therapy. Randomized, double-blind controlled clinical trials are the gold standard of study designs. On the numeric scale of study design, randomized controlled trials, as well as meta-analyses of randomized controlled trials, provide level 1 evidence. Non-randomized controlled trials and subset analyses of randomized controlled trials provide level 2 evidence, while cohort studies, case series and case-control studies are considered level 3 evidence (32).

The strength of study endpoints is designated by an alphabetical scale. Total mortality is the strongest endpoint (A), as it is the most easily defined and least subject to investigator bias. Cause-specific mortality is level B as it is more subjective, being prone to investigator bias and possibly missing treatment effects that can reduce overall survival. Quality-of-life assessments are considered level C endpoints. Indirect surrogate endpoints, such as event-free survival, disease-free survival, progression-free survival and tumor response rate, are level D due to the potential subjectivity of investigator interpretation. These latter endpoints do not necessarily translate into an improved overall survival or quality-of-life, but have particular value due to the shorter course in which they can be achieved. (32)

The NCCN offers an alternate scale for categories of evidence and consensus that reflects the strength of a treatment recommendation. Category 1 evidence and consensus is based upon a high-level of evidence with uniform NCCN consensus that the intervention is appropriate. Category 2A is based upon lower levels of evidence, but there is still uniform consensus that the intervention is appropriate. Category 2B reflects a lower level of evidence with consensus that the intervention is appropriate, although not necessarily uniform consensus. Category 3 is based upon any level of evidence and indicates major NCCN disagreement that the intervention is appropriate (33).

CONCLUSION

Cancer therapies generate a substantial volume of clinical trials and evidence. An understanding of the principles of clinical trials and the value of their outcomes is necessary to critically assess their applicability to one's own patients, and represents the crucial first step in providing appropriate and optimal cancer care.

Acknowledgments

Funding: RS is supported in part by NIH CA126809.

References

1. U.S. Food and Drug Administration. Available from: <http://www.fda.gov/Drugs/ResourcesForYou/Consumers/default.htm>
2. Miller AB, Hoogstraten B, Staquet M, Winkler A. Reporting results of cancer treatment. *Cancer*. 1981 Jan 1; 47(1):207–214. [PubMed: 7459811]
3. Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer*. 2009 Jan; 45(2):228–247. [PubMed: 19097774]
4. Odgaard-Jensen J, Vist GE, Timmer A, et al. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev*. (4):MR000012.
5. Friedman, LM.; Furberg, C.; DeMets, DL. *Fundamentals of clinical trials*. 4. New York: Springer;
6. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ*. 1999 Sep 11; 319(7211):670–674. [PubMed: 10480822]
7. Therasse P, Arbuck SG, Eisenhauer EA, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *Journal of the National Cancer Institute*. 2000 Feb 2; 92(3):205–216. [PubMed: 10655437]
8. Forner A, Ayuso C, Varela M, et al. Evaluation of tumor response after locoregional therapies in hepatocellular carcinoma: are response evaluation criteria in solid tumors reliable? *Cancer*. 2009 Feb 1; 115(3):616–623. [PubMed: 19117042]
9. Bruix J, Sherman M, Llovet JM, et al. Clinical management of hepatocellular carcinoma. Conclusions of the Barcelona-2000 EASL conference. European Association for the Study of the Liver. *J Hepatol*. 2001 Sep; 35(3):421–430. [PubMed: 11592607]
10. Lencioni R, Llovet JM. Modified RECIST (mRECIST) Assessment for Hepatocellular Carcinoma. *Semin Liver Dis*. Feb; 30(1):52–60. [PubMed: 20175033]
11. Lu TL, Becce F, Bize P, et al. Assessment of liver tumor response by high-field (3 T) MRI after radiofrequency ablation: short- and mid-term evolution of diffusion parameters within the ablation zone. *Eur J Radiol*. Sep; 81(9):e944–950. [PubMed: 22817977]
12. Rhee TK, Naik NK, Deng J, et al. Tumor response after yttrium-90 radioembolization for hepatocellular carcinoma: comparison of diffusion-weighted functional MR imaging with anatomic MR imaging. *J Vasc Interv Radiol*. 2008 Aug; 19(8):1180–1186. [PubMed: 18656011]
13. Yuan Z, Ye XD, Dong S, et al. Role of magnetic resonance diffusion-weighted imaging in evaluating response after chemoembolization of hepatocellular carcinoma. *Eur J Radiol*. Jul; 75(1):e9–14. [PubMed: 19540083]
14. Sahin H, Harman M, Cinar C, et al. Evaluation of treatment response of chemoembolization in hepatocellular carcinoma with diffusion-weighted imaging on 3.0-T MR imaging. *J Vasc Interv Radiol*. Feb; 23(2):241–247. [PubMed: 22019180]
15. Mannelli L, Kim S, Hajdu CH, et al. Assessment of tumor necrosis of hepatocellular carcinoma after chemoembolization: diffusion-weighted and contrast-enhanced MRI with histopathologic correlation of the explanted liver. *AJR Am J Roentgenol*. 2009 Oct; 193(4):1044–1052. [PubMed: 19770328]

16. Wahl RL, Jacene H, Kasamon Y, Lodge MA. From RECIST to PERCIST: Evolving Considerations for PET response criteria in solid tumors. *J Nucl Med*. 2009 May; 50(Suppl 1): 122S–150S. [PubMed: 19403881]
17. Streiner DL. The 2 “Es” of research: efficacy and effectiveness trials. *Can J Psychiatry*. 2002 Aug; 47(6):552–556. [PubMed: 12211883]
18. Pazdur R. Endpoints for assessing drug activity in clinical trials. *Oncologist*. 2008; 13(Suppl 2): 19–21. [PubMed: 18434634]
19. Bergmann L, Berns B, Dalglish AG, et al. Investigator-initiated trials of targeted oncology agents: why independent research is at risk? *Ann Oncol*. Aug; 21(8):1573–1578. [PubMed: 20133383]
20. Buyse M, Sargent DJ, Saad ED. Survival is not a good outcome for randomized trials with effective subsequent therapies. *J Clin Oncol*. Dec 10; 29(35):4719–4720. author reply 4720–4711. [PubMed: 22042965]
21. Gill S, Sargent D. End points for adjuvant therapy trials: has the time come to accept disease-free survival as a surrogate end point for overall survival? *Oncologist*. 2006 Jun; 11(6):624–629. [PubMed: 16794241]
22. Zhuang SH, Xiu L, Elsayed YA. Overall survival: a gold standard in search of a surrogate: the value of progression-free survival and time to progression as end points of drug efficacy. *Cancer J*. 2009 Sep-Oct; 15(5):395–400. [PubMed: 19826359]
23. Di Leo A, Buyse M, Bleiberg H. Is overall survival a realistic primary end point in advanced colorectal cancer studies? A critical assessment based on four clinical trials comparing fluorouracil plus leucovorin with the same treatment combined either with oxaliplatin or with CPT-11. *Ann Oncol*. 2004 Apr; 15(4):545–549. [PubMed: 15033657]
24. Llovet JM, Di Bisceglie AM, Bruix J, et al. Design and endpoints of clinical trials in hepatocellular carcinoma. *J Natl Cancer Inst*. 2008 May 21; 100(10):698–711. [PubMed: 18477802]
25. Llovet JM, Ricci S, Mazzaferro V, et al. Sorafenib in advanced hepatocellular carcinoma. *N Engl J Med*. 2008 Jul 24; 359(4):378–390. [PubMed: 18650514]
26. Cheng AL, Kang YK, Chen Z, et al. Efficacy and safety of sorafenib in patients in the Asia-Pacific region with advanced hepatocellular carcinoma: a phase III randomised, double-blind, placebo-controlled trial. *Lancet Oncol*. 2009 Jan; 10(1):25–34. [PubMed: 19095497]
27. Dawson, B.; Trapp, RG. Basic & clinical biostatistics. 4. New York: Lange Medical Books/McGraw-Hill, Medical Pub. Division; 2004.
28. Zwillinger, D.; Kokoska, S. CRC standard probability and statistics tables and formulae. Boca Raton: Chapman & Hall/CRC; 2000.
29. Moye LA. P-value interpretation and alpha allocation in clinical trials. *Ann Epidemiol*. 1998 Aug; 8(6):351–357. [PubMed: 9708870]
30. Freiman JA, Chalmers TC, Smith H Jr, Kuebler RR. The importance of beta, the type II error and sample size in the design and interpretation of the randomized control trial. Survey of 71 “negative” trials. *N Engl J Med*. 1978 Sep 28; 299(13):690–694. [PubMed: 355881]
31. Bland JM, Altman DG. One and two sided tests of significance. *BMJ*. 1994 Jul 23.309(6949):248. [PubMed: 8069143]
32. National Cancer Institute Levels of Evidence for Adult and Pediatric Cancer Treatment Studies (PDQ®). [cited 2012 November]; Available from: <http://www.cancer.gov/cancertopics/pdq/levels-evidence-adult-treatment/HealthProfessional/page1>
33. National Comprehensive Cancer Network Categories of Evidence and Consensus. Available from: http://www.nccn.org/professionals/physician_gls/categories_of_consensus.asp

Table 1

Comparison of WHO, RECIST 1.0, RECIST 1.1 and mRECIST

	WHO	RECIST 1.0	RECIST 1.1	EASL (HCC)	mRECIST (HCC)
Complete Response (CR)	Disappearance of all lesions	Disappearance of all lesions	Disappearance of all lesions and pathologic lymph nodes	Disappearance of all lesions	Disappearance of any intratumoral enhancement in all target lesions
Partial Response (PR)	50% decrease in sum of cross-products (longest diameter multiplied by longest perpendicular diameter)	30% decrease in the sum of the longest diameters	30% decrease in the sum of the longest diameters	50% decrease in sum of cross-products of enhancing lesion	30% decrease in the sum of the longest diameters of viable (enhancement in arterial phase) target lesions
Stable Disease (SD)	Does not meet criteria for CR or PR	Does not meet criteria for CR or PR	Does not meet criteria for CR or PR	Does not meet criteria for CR or PR	Does not meet criteria for CR or PR
Progressive Disease (PD)	25% increase in one or more measurable lesion or appearance of a new lesion	20% increase in the smallest sum measured on any prior study, or appearance of new lesions	20% increase in the smallest sum measured on any prior study but increase must be at least 5mm, or appearance of new lesions	25% increase in enhancing lesion or appearance of new lesion	20% increase in the smallest sum of the diameters of viable (enhancement in arterial phase) target lesions measured on any prior study
Definition of Measurable Disease	No limitation on minimal size	10 mm on spiral CT 20 mm on non-spiral CT	10 mm (no distinction between spiral/non-spiral)		
Number of Lesions to Measure	Not specified	10 lesions total (5 per organ)	5 lesions total (2 per organ)		
Method of Measurement	Cross-product: longest diameter multiplied by longest perpendicular	Longest diameter in axial plane	Longest diameter in axial plane	Cross-product: longest diameter multiplied by longest perpendicular of enhancing lesion	Longest diameter of viable target lesions in axial plane
Lymph Node	No mention	No mention	15 mm short axis for target <10 mm is not pathological	No mention	Porta hepatitis lymph node considered malignant if the short axis >20 mm
New/Worsening Ascites or Pleural Effusion	No mention	Cytologic confirmation required*	Cytologic confirmation required*	No mention	Cytologic confirmation required*

* Cytologic confirmation of the neoplastic nature of any effusion that appears or worsens during treatment is required when the measurable tumor has met criteria for response or stable disease. Under such circumstances, the cytologic examination of the fluid collected will permit differentiation between response and stable disease